

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.
OM protein - protein search, using sw model
Run on: July 30, 2002, 08:10:43 ; Search time 71.68 Seconds
(without alignments)
715.844 Million cell updates/sec
Title: US-09-836-461-2
Perfect score: 2785
Sequence: 1 MRVLCAPPEAMPSSNSRPPA.....PVTMGFFVKNVNALACRYR 534
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match %	Length DB ID	Description
1	2478	89.0	480 2	JC7506 heparanase protein
2	302	10.8	521 2	T45608 hypothetical prote
3	155	5.6	190 2	T01953 hypothetical prote
4	111	4.0	1260 2	T14022 reverse transcript
5	109	3.9	986 2	F98229 sarcosine oxidase
6	109	3.9	986 2	AH3056 sarcosine oxidase
7	107	3.8	2026 1	OYBY adenylate cyclase
8	105	3.8	1020 2	T18342 glutamate dehydrog
9	104.5	3.8	510 2	T44810 conserved hypothet
10	103	3.7	763 2	H84432 probable helicase
11	101.5	3.6	629 2	C64180 hypothetical prote
12	101.5	3.6	1125 1	JH0771 protein-tyrosine k
13	101	3.6	535 2	S58740 cytochrome-c oxida
14	100.5	3.6	582 2	A84206 hypothetical prote
15	100.5	3.6	630 2	C71374 probable glucose i
16	99.5	3.6	455 2	S67627 probable membrane
17	99.5	3.6	1272 2	C90593 hypothetical prote
18	99	3.6	2403 2	T30875 PRP8 protein homol
19	98.5	3.5	335 2	E84992 tryptophan--trna 1
20	98	3.5	845 1	G72079 endopeptidase Clp
21	98	3.5	845 2	C86545 ClpC proteinase [i
22	98	3.5	845 2	G81590 ATP-dependent Clp
23	97.5	3.5	1517 2	B81393 DNA-directed RNA p
24	96.5	3.5	969 2	D82895 DNA polymerase III
25	96	3.4	832 2	E71492 hypothetical prote
26	96	3.4	898 2	B84471 cytoplasmic acotin
27	96	3.4	1229 2	S42391 SIP3 protein - yea
28	96	3.4	4367 1	B54802 dynein heavy chain
29	95.5	3.4	406 2	S68866 qsoA protein - Co

ALIGNMENTS

RESULT 1
JC7506
heparanase protein 2a - human
C:Species: Homo sapiens (man)
C:Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 01-Dec-2000
C:Accession: JC7506
R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hirccock, M.;
Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000 - 04-5
A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase famil
A:Reference number: JC7506
A:Accession: JC7506
A:Molecule type: mRNA
A:Residues: 1-480 <MCK>
A:Cross-references: GB:AF282885
C:Comment: This protein, a intracellular membrane-bound enzyme, has biological and th
therapies.
C:Genetics:
A:Gene: hpa2a
A:Map position: 10q23-10q24
C:Keywords: heparin binding; membrane bound

Query Match		89.0%;	Score 2478;	DB 2;	Length 480;
Best Local Similarity		89.9%;	Pred. No. 2e-184;		
Matches 480;		Conservative	0;	Mismatches	0;
				Indels	54;
				Gaps	1;
Qy	1	MRVLCAPPEAMPSSNSRPPACIAPGALYIALLHLSSQAGDRRPLPVDRAAGLKEKTL	60		
Db	1	MRVLCAPPEAMPSSNSRPPACIAPGALYIALLHLSSQAGDRRPLPVDRAAGLKEKTL	60		
Qy	61	ILLDYSTKNPVTNVNENFLSLQDPSIIHDGWLDFLSSKRLVTLARGLSPAFLRFGGKRT	120		
Db	61	ILLDYSTKNPVTNVNENFLSLQDPSIIHDGWLDFLSSKRLVTLARGLSPAFLRFGGKRT	120		
Qy	121	DFLQFNLRNPAKSRGGGPDYILKNYEDDIVRSDVALDKQKCKIAQHPDVMLELQREK	180		
Db	121	DFLQFNLRNPAKSRGGGPDYILKNYEDDIVRSDVALDKQKCKIAQHPDVMLELQREK	180		
Qy	181	AAQMHVLVLLKEQFSNTYSNLIITPENNYRTMHGRAVNGSQLGKDYIQLKSLQPIRIYSR	240		
Db	181	AAQMHVLVLLKEQFSNTYSNLIITPENNYRTMHGRAVNGSQLGKDYIQLKSLQPIRIYSR	240		
Qy	241	ASLYGNIGRPRKNVITALLDGFEMKVAGSVTDATVWQHCVYIDGRVVKVMDFLKRLDITLS	300		
Db	241	ASLYGNIGRPRKNVITALLDGFEMKVAGSVTDATVWQHCVYIDGRVVKVMDFLKRLDITLS	300		
Qy	301	DOIRIKQVNVNYPGCKITWLEGVVTTSAGGTNNISDSYAAGFLWLTGMLANQIDVV	360		
Db	301	DOIRIKQVNVNYPGCKITWLEGVVTTSAGGTNNISDSYAAGFLWLTGMLANQIDVV	360		
Qy	361	IRHSFFDHGYNHLVDQNFNPLPDYWLISLYIKRLIGPKVLAVHVGLOKRPGRVIRDKL	420		
Db	361	IRHSFFDHGYNHLVDQNFNPLPDYWLISLYIKRLIGPKVLAVHVGLOKRPGRVIRDKL	420		

Db 307 IRHSFFDHGYNHLVDQNFDPYWLSSLLYKRLIGPKVLAVHVGAGLQKPRGVRDKL 366
QY 421 RIYACTNHHNNHYRGSTITLFIINLHRSRKKIKLAGTLRDKLVHQLQYGGQGLSK 480
Db 367 RIYACTNHHNNHYRGSTITLFIINLHRSRKKIKLAGTLRDKLVHQLQYGGQGLSK 426
QY 481 SVOLNGQPLVMVDDGTLPELKPRLRAGRTLVIPTVMGFFVVKNNALACRYR 534
Db 427 SVOLNGQPLVMVDDGTLPELKPRLRAGRTLVIPTVMGFFVVKNNALACRYR 480

RESULT 2
T45608
hypothetical protein F13G24.30 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T45608
R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba
submitted to the Protein Sequence Database, December 1999
A:Reference number: 223009
A:Accession: T45608
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-521 <BE>
A:Cross-references: EMBL:AL133421
A:Experimental source: cultivar Columbia; BAC clone F13G24
C:Genetics:
A:Map position: 5
A:Introns: 53/3; 66/1; 127/2; 177/1; 256/1; 319/2; 361/2; 394/3
A:Note: F13G24.30

Query Match 10.8%; Score 302; DB 2; Length 521;
Best Local Similarity 24.8%; Pred. No. 2e-15;
Matches 139; Conservative 73; Mismatches 183; Indels 166; Gaps 27;

QY 75 NENFLSLQID--PSIHD-----GWLDFLS---SKRLVTLA-RGLSPAFLRFGGR 119
Db 22 DENFVCATLDWNP---HDKCNYDCWPYSSVINMDLTRPLTKAIKAFKPLRIRIGSL 78
QY 120 TDFLQFQ--NLNRP-----KSRGGPGDYLYLNKYEDDVRSDVALDKQKGLAQHPDM 173
Db 79 QDOVIYDVGNLKTCPRPFOKMNSG-----LFGFSKGLHMKRWD-- 117
QY 174 LEIQREKAAQMHVL-----LLKEQFSNTYSNLILTEPNRYRMHGRAVN--- 217
Db 118 -ELNSFLTATGAVTTGLNALRGRHKLRKANGGAWDHINTQDFLNTYVSKGYVIDSWEF 176
QY 218 GSOL-----GKDYQLKSLQPIRIYSRASLYGPNIGRPRKNVIALDGF- 262
Db 177 GNELSGSGVGSVSAELYGKDLIVLKDVN--KVYKNSMLHKPILVAP-----GGFY 226
QY 263 -----MKVAG-STVDATVWQHCHIDGR-----VWVMDFLKTRLDTLSDQIRKIQ 307
Db 227 EQOYTKLLEISGPSVDVVT-HHIYNLGSGNDPALVKXIMD---PSYLSQVSKTFKQVN 282
QY 308 KVYNTVTPGKKIWLEGWTTTSAG-----GTNNLSDSAAGFLWLNTLGLANOGIDVVIH 363
Db 283 QTQEHGP-----WASFWGESGAYNSGGRHVSDFIDSEFWDQLGMSARHNTKYVCQ 338
QY 364 SFFDHGYNHLVDQNFDPYWLSSLLYKRLIGPKVLAVHVGAGLQKPRGVRDKLRY 423
Db 339 TLVGGFYGLEKCTFVPNPDPYSALLHRLMGKGLAVQTDG-----PPQLRVY 387
QY 424 AHCNTNHHNNHYRGSTITLFIINLH-----RSRKKIKLACTLRDKLV 464
Db 388 AHCSKG-----RAGVTLLINLSNQSDFTVSVNSGINVVLNAESRKKSLDITLRKPS 441
QY 465 -----HOYLLQPYGEG-LKSKSVOLNGQPLVMVDDGTLPELKPRLRAGRT 510
Db 442 WIGSKASDGYLNREYHLTP--ENGVLRSKTYWNLGKSLKPTATGDIPLSVLRVSNP 499
QY 511 LVIPPTVMGFFVVKNNALAC 531

Db 500 LNVLPLSMSFIVLPNFASAC 520

RESULT 3
T01953
hypothetical protein T2L5.6 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 21-Jan-2000
C:Accession: T01953
R:Geisel, C.; Smith, A.; Le, T.
submitted to the EMBL Data Library, October 1998
A:Description: The sequence of A. thaliana T2L5.
A:Reference number: Z14470
A:Accession: T01953
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-190 <GEI>
A:Cross-references: EMBL:AF096371; NID:g3695386; PID:g3695392
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:Introns: 36/2; 69/3
A:Note: T2L5.6
C:Superfamily: Arabidopsis thaliana hypothetical protein T2L5.6

Query Match 5.6%; Score 155; DB 2; Length 190;
Best Local Similarity 24.9%; Pred. No. 0.00011;
Matches 49; Conservative 32; Mismatches 70; Indels 46; Gaps 6;

QY 362 RHSEFFDHGYNHLVDQNFNPDPYWLSSLLYKRLIGPKVLAVHVGAGLQKPRGVRDKLR 421
Db 12 RQSLGGNGYGLLNTNFTPNPDYYSALIWQLMGRKALFTTFSGTK-----KIR 60
QY 422 IYAHCTNHHNNHYRGSTITLFIINLHRSR---KTKLAGTLRDKLVHQLQY----- 472
Db 61 SYTHCAROSK-----GITVLLMLNDLTTTVVAKVELNNSF---SLRHTKHKMSYKRASSQ 112
QY 473 ---GOEG-----LKSKSVOLNGQPLVMVDDGTLPELKPRLRAGRTLVI 514
Db 113 LFGPNGVIOREYHLTKADGNLHQSMTLLNGNALQVNSMGDLPPPIEHINSTEPIIA 172
QY 515 PVTMGFFVVKNNALAC 531
Db 173 PYSIVFVHMNRNVVPAC 189

RESULT 4
T14022
reverse transcriptase homolog - slime mold (Dictyostellium discoideum) retrotransposab
C:Species: Dictyostellium discoideum
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C:Accession: T14022
R:Winckler, T.; Tschepke, C.; de Hostos, E.L.; Jendretzke, A.; Dingermann, T.
Mol. Gen. Genet. 257, 655-661, 1998
A:Title: Tdd-3, a transfer RNA gene-associated poly(A) retrotransposon from Dictyoste
A:Reference number: Z17858; MUID:98265925
A:Accession: T14022
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1260 <WTN>
A:Cross-references: EMBL:AF002669; NID:g2558820; PID:g2558822; PIDN:AAC48324.1

Query Match 4.0%; Score 111; DB 2; Length 1260;
Best Local Similarity 19.7%; Pred. No. 5.2; Mismatches 89; Indels 232; Gaps 34;

QY 69 NPVRTVWNEFLSLQIDPPSIHDGWL---DFL---SKRLV-----TLARGLSP 110
Db 94 NGIGILNHNQNKLSPIIEGRLLISDILIKDTTTRILAIYAPAPQDKRKTASTLNK 153

Db 362 DQRLDEARALGIEVLAGHSVYTAGRLRVSSMTVGRNGSGNKRKKAIDALVVSAGW--- 418
QY 96 LSSKRLVTLARGLSPAFLEGRKRTDFLOFQNLRN-----PAKSRGG 137
Db 419 TPSVHLFSQSRGK----LKFDAAQRFLPDHIVQNCVSVSGACNGTDLDVAITAEAAAGG 474
QY 138 PGDPYILKN-----YEDDIVRSDVALDKQKCKIAQHFPD 171
Db 475 GSVAFSGENARAWTGMGIAAGAGGTGVKAFIDQHDVCAKDIRLAVREG----- 526
QY 172 VMLELQREKAAQHMLVLLKKEQFNTSYNLLITPEPNRYRTHWGRAVNSQLGKDY--IQLK 229
Db 527 -----MHSV---EHIKRFTTNGMASDQKSKMNMHGLAIASEALGRDLPRVGLT 571
QY 230 SLLQP-----IRIYSRASLYGPNIGRPKNVIALLDG-----EMKVAGS 268
Db 572 TFRQPTPTVFTGLINHSRGALEDPTRKTPMHKEE-ELAAGAVFEDVGNKRAWFFPRAGE 530
QY 269 TVDAVTHQHCYIDGRVVKVNDFLKTRLLDLSQIRKIQKV-----NTYTPGKKIWLLEG 323
Db 631 DMHEAINRECKTVRTSGVGFEDASTLKGIEVVGPDAAKFLNLIVTNAWDTLKPGRCRY--G 688
QY 324 VVITSAGTNLSDSYAAGFLWLNTLGLMLANOGIDVVIHRSFEDHGYNHLVDQNFPLPD 383
Db 689 IMTREGFV-----YDDG-----VVGRLAEDRFHTTTGGAPRVLOHMEYLTQTEFPD 737
QY 384 --YWLSLLYKR-----LIGPK-----VLAVHVAGLQKRP-----PGRV--- 415
Db 738 LNVWLTSATEQMAVIAVQGPAREVIAFPVEGIDLSPEAFPHMAVAEGKFCGVPRLFRV 797
QY 416 -----IRDKLRIYACHTNHHNHNVRGSIITLFIINLH----- 447
Db 798 SFTGELGFEINVPADYGAAVSAIRDRTAEGVC-----LYGTETMHLRAEKGYII 849
QY 448 -----RSRKKIKLAGTLRDKLVHVOYLLQPYGOEGL 477
Db 850 VGQDTDGTVPDDAGLAWAVSKKKTDFVGIKRLKIDLTTRTGKQLV-----GL 898
QY 478 KSKSVOLNQPLVMVDG-----TLPLKPRPL----- 505
Db 899 KTK-----DRLTVPDEGGQIVTDPN-QPKPMTMLGHVTSAYWSENGLHGSIAFALVADGR 951
QY 506 -RAGRTLVP 514
Db 952 ARMGETLYLP 961
RESULT 7
OVBX
adenylate cyclase (EC 4.6.1.1) - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: ATP pyrophosphate-lyase; protein J1401; protein YJL005w
C:Species: *Saccharomyces cerevisiae*
C>Date: 28-Dec-1987 #sequence_revision 08-Sep-1995 #text_change 21-Jan-2000
C:Accession: S56776; S56775; A24776; S05828; S55183
R:To Van, D.; Perea, J.; Jacq, C.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S56776
A:Accession: S56776
A:Molecule type: DNA
A:Residues: 1-1823 <DEH>
R:de Haan, M.; Grivell, L.A.; Smits, P.H.M.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S56771
A:Accession: S56775
A:Molecule type: DNA
A:Residues: 673-2026 <ZAG>
R:Kataoka, T.; Broek, D.; Wagler, M.
Cell 43, 493-505, 1985
A:Title: DNA sequence and characterization of the *S. cerevisiae* gene encoding adenylate
A:Reference number: A24776; MUID:86079531
A:Accession: A24776

A:Molecule type: DNA
A:Residues: 1-261, 'L', 263-547, 'L', 549-591, 'H', 593-708, 'I', 710-961, 'P', 963-1387, 'S', 13
A:Cross-references: EMBL:M12057; NID:g171359; PIDN:AAA34549.1; PID:g171360
A:Note: the authors translated the codon TTA for residue 262 as Ser, ACG for residue
2 as Leu, TCA for residue 1566 as Ala, GAG for residue 1659 as Ala, GGT for residue 1
R:Masson, P.; Lenzen, G.; Jacquemin, J.M.; Danchin, A.
Curr. Genet. 10, 343-352, 1986
A:Title: Yeast adenylate cyclase catalytic domain is carboxy terminal.
A:Reference number: S05828; MUID:88165073
A:Accession: S05828
A:Molecule type: DNA
A:Residues: 1042-1426, 'D', 1428-1460, 'T', 1462-1955, 'V', 1957-2008, 'M', 2014-2015, 'TNEFKV
A:Cross-references: EMBL:X03449; NID:g3487; PIDN:CAA27175.1; PID:g34488
R:de Haan, M.; Smits, P.H.M.; Grivell, L.A.
submitted to the EMBL Data Library, May 1995
A:Reference number: S55183
A:Accession: S55183
A:Molecule type: DNA
A:Residues: 673-2026 <DEW>
A:Cross-references: EMBL:X87611; NID:g854567; PIDN:CAA60917.1; PID:g854568
C:Genetics:
A:Gene: SGD:CYR1; CDC35; MIPS:YJL005w
A:Cross-references: SGD:S0003542; MIPS:YJL005w
A:Map position: 10L
C:Superfamily: yeast adenylate cyclase; leucine-rich alpha-2-glycoprotein repeat homo
C:Keywords: CAMP biosynthesis; duplication; phosphorus-oxygen lyase; tandem repeat
F:669-1343/Region: leucine-rich 23-residue repeats
F:1085-1087/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR>
F:1610-2026/Domain: catalytic #status predicted <CAT>
F:1664-1749/Domain: yeast adenylate cyclase catalytic domain homology <YACC>

Query Match 3.8%; Score 107; DB 1; Length 2026;
Best Local Similarity 20.2%; Pred. No. 22;
Matches 124; Conservative 82; Mismatches 207; Indels 200; Gaps 31;
QY 37 LSSQAGRRPLPVDRAAGLKE-KTLLLDVSTKNPV---RTVNFNLSLQDLPST----- 87
Db 891 LNLQCNELSLP---AGFVELKNLQLLDSSNKPMPHYEVINYCTNLLQIDLSYNKIQS 946
QY 88 -----IHGWLDFLSKKRLVTLARGLSPAFLEGRKRTDFLOFQNL----- 128
Db 947 LPQSTKYLVKAKMNLSHNKLNFIGDLSMTDLRTLNLRNRISSIKTNASNLQNLFLTD 1006
QY 129 -----RNPAGSRGGPGDYIYLNKVEDDIVRSDVALDKQKCKIAQH 169
Db 1007 NRISNFEDTLKRLALETOENPITSIS--FKDFYPKN-----WTSLLTNK--AQLSSI 1055
QY 170 PDVMLE----LOREKAAQHMLVLLKKEQFNTSYNLLITPEPN-----YRT- 210
Db 1056 PGEILLTKLSFLEKLELNQNNLTRLPQETSKLTKLVFLSVARNKLYIPPELSQLSLRTL 1115
QY 211 -MHGRAVNSQLGKDYIOLKSLQPIRIYSRASL-----YCPNIGRPKNVIALLD 260
Db 1116 DLHSNNIRDFVDGMENLELTSLNISSNAGSSLENSFHNMSYGSKLSK-----SLM- 1168
QY 261 GFMKVAGSTVDVATWQ--HCYIDGRVVKVMDFLKTRLLDLSQIRKIQKVNTYTPGKK 318
Db 1169 -FFTAADNQDQDAMWPLENCFVN---LKVNLSTVNFSDV---SHMKLESTELYLSNKK 1221
QY 319 IWLEGVVTSAGGTNNLSDSYAAGFLW--LNTLGLMANQGDIV-----VIHRSFFDHGY 370
Db 1222 L-----TTLSGDT-----VLKWSLTKTLMNSNQMLSLPAELNSLSQLSVDFVGA 1266
QY 371 NHLVDQNFNPLPDY-WLS---LLYKRLIGP-----KVLA 400
Db 1267 NQKYNISNYHYDWNWRNKKELKYNLNSGNRRFEIKPSIHDDADLSLTVLPOLKVLG 1326
QY 401 VHVAGLQKRPGRVIRDKLRIYACHTNHHNH-----NYV-----RGSIITL 441
Db 1327 LMDVTLNTTKVPDENVNFRLRTTASIINGMRVYGVADTLGQRDYVSSRDVTFERFRGNDE 1386
QY 442 FIINLHRSRKK-----IKLAGTLRDK-LVHQVLLQPYGOE-----GLSKSVOLN 485

Db 1387 CLCLHDSKKNADYGHNISRVIRDIYDKILIRQ--LERYGDETDDNTKTALRFSFLQLN 1444
Qy 486 GQ---PLVMVDDG 495
Db 1445 KEINGMLNSVDNG 1457
RESULT 8
T18342
glutamate dehydrogenase (EC 1.4.1.2) precursor - Sauroleishmania tarentolae
C:Species: Sauroleishmania tarentolae
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-May-2000
C:Accession: T18342
R:Bringingaud, F.; Srippecke, R.; Frech, G.C.; Freedland, S.; Turck, C.; Byrne, E.M.; Simps
Mol. Cell. Biol. 17, 3915-3923, 1997
A:Title: Mitochondrial glutamate dehydrogenase from Leishmania tarentolae is a guide RNA
A:Reference number: 218878; MUID:97342629
A:Accession: T18342
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1020 <BRI>
A:Cross-references: EMBL:U31177; NID:9945092; PID:9945093; PIDN:AAB62735.1
C:Genetics:
A:Gene: GDH
C:Superfamily: glutamate dehydrogenase
C:Keywords: NAD; oxidoreductase

Query Match 3.8%; Score 105; DB 2; Length 1020;
Best Local Similarity 25.5%; Pred. No. 11;
Matches 42; Conservative 30; Mismatches 51; Indels 42; Gaps 7;
Qy 114 RFGKRTDFLQFLRNPAKSRGGP---GPDYILKNYEDDIVRSDVA-----LDKQ 161
Db 656 KGLKSEMRKFTQ-----GGPDGLSGNEVLRSKKKWGVWDISASLDHPNGIDRE 707
Qy 162 KGCKIAQHPDVMLEQRKAQMHLLVLEKQFSNTYSNLITPEPNYRTMHGRAV-NGSQ 220
Db 708 ELARLAHRLPLREFRSK-----LSPEGF-----LVLTEDHNVLKLPDGTTLVEDGSR 754
Qy 221 LGGDYIOLKSLLOPIRIYRSASLYGPNIGRPRKNVIALLDGFMKV 265
Db 755 LRNEHFLK-----YSDADVFVPCGGRPRSVTLENVGRFLKV 791

RESULT 9
T44810
conserved hypothetical protein yngK [imported] - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 28-Jul-2000
C:Accession: T44810
R:Duitman, E.H.; Hamoen, L.W.; Rembold, M.; Venema, G.; Seitz, H.; Saenger, W.; Bernhar
Proc. Natl. Acad. Sci. U.S.A. 96, 13294-13299, 1999
A:Title: The mycosubtilin synthetase of Bacillus subtilis ATCC6633: A multifunctional hy
A:Reference number: 222848; MUID:20027541
A:Accession: T44810
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-510 <DUI>
A:Cross-references: EMBL:AF184956; NID:96449053; PIDN:AAF08799.1; PID:96449059
A:Experimental source: strain ATCC6633
C:Genetics:
A:Gene: yngK
C:Superfamily: conserved hypothetical protein yngK

Query Match 3.8%; Score 104.5; DB 2; Length 510;
Best Local Similarity 19.1%; Pred. No. 4.2;
Matches 100; Conservative 87; Mismatches 196; Indels 141; Gaps 25;
Qy 3 VLCAFPPEAMP-SSNSRPPACLPAGALYLLALLHLHLSLSSQAGDRRLPVPDRAGLKEKTLI 61

Db 16 IIGTETVISPVFMANAQPRELR--AVWIASVNLIDWPSKKG-----LPVEEQ---KOEYIK 66
Qy 62 LLDVSTKNPVRTNENFLSLQLDPSI-----IHGWLDLFLSSKRLVTLARGLSPAF--L 113
Db 67 LLD-----DVQKMGMAVIVQIKPTADAFPSAYGPWSEYL-----TCVQKDKGYPDL 115
Qy 114 RFGKRTD--FLOFONLRNPAKSRGGPDYLYKNYEDDIVRSDVALDK-QKGCKIAQHP 170
Db 116 AFMIEETHKRNLEPHAFWPNYR-----ITWNHTDLNKLKSKDHPARQHP 158
Qy 171 DVMLELQREKRAQMHLLVLLKEQFSNTYSNLITPEPNYRTMHGRAVNSQLGKDYIQLKS 230
Db 159 D-----WVAAYGNQLYHP-----GISEARDFI-VKG 184
Qy 231 LIQPIRIYRSASLYGPNIGRPRKNVIALLDGFMKVAGSTVDVATWQHCVIDGRVVKVMDF 290
Db 185 IEEVVKHYDIDAVHMDDFYF-----YKTAGEFFPDHAQYEQYKGTFTFSNIDWW 233
Qy 291 LKTRLLDLSDOIIRKIQKVNTYTPGKKIWL- -GVVTTSA-----GCTNNLSDSY 339
Db 234 RR----DNVNQLVKQINQTIKTAKPVKFGISPGFVGRNAADDPGTSNTKAGVRNYDDLY 289
Qy 340 RAGFLWLNTLGLMANOGIDVIRHSFFDHGYNHLYVDQNFNPLDYWLSLLYKRLIGPKVL 399
Db 290 ADTRHWIQ-----EGDIDYITAPQIYWSIGFNAAA---YDVLADWMSNEVKNR---PVHL 337
Qy 400 AVHVAGLQKRPGRVIRDKLRIYAHCTNHHNHNVVRSITLFIINLHRSRKKIKLAGTL 459
Db 338 YIGQAYKINNFDPPWSDPEYVVRQITLNRQLELVKSGMHFSLKDLNKNPFGIK----- 392
Qy 460 RDKLVHQYLLQPGYQEGILKSKSVQLNGQPLVMVDDGTLPELKPR 503
Db 393 -DRLITE-----LYSKPALVPQMPWL---DSTAPK-KPK 421

RESULT 10
H84432
probable helicase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: H84432
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: H84432
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-763 <STO>
A:Cross-references: GB:AE002093; NID:g4038042; PIDN:AAC97224.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g02090
A:Map position: 2

Query Match 3.7%; Score 103; DB 2; Length 763;
Best Local Similarity 21.6%; Pred. No. 10;
Matches 100; Conservative 56; Mismatches 133; Indels 174; Gaps 25;
Qy 13 SSNSRPPACLPAGALYLLALLHLHLSL-----SSQAGDRRLP-----VDRAAGLKERT 59
Db 310 SKAGKPPP-----FNVLLVCYSLFERHSEQQKDRKLVKRWKWSVCLMDENHALKDN 362
Qy 60 LI-----LLDVSTKNPVRTNENFL-----SLQDPSIHDGW--LDLSSKRLVTLARGLS 109
Db 363 SYRWKNLMSVA-----RANQRMLTGTPLQND---LHELWLLLEFMLPDIET----- 408
Qy 110 PAFLRFGKRTDFLQFN-----LRNPAKSRGGPDYLYKNYEDDIVRSDVALDKQKCC 164
Db 409 -----ENVDKLLNAEDTELITRMKSILGP---FTLRRLKSDVMQQLVP----- 450

	Query Match	3.6%	Score 100.5;	DB 2;	Length 582;	
	Best Local Similarity	21.2%;	Pred. No. 10;			
	Matches	69;	Conservative	49;	Mismatches	110; Indels 97; Gaps 14;
<hr/>						
Qy	49 VDRAAGLKEKTLILLDVS	TKNPVTNENFLSLQLDSP	IIHDGWLDFLSKKRLVTL	LARGL	108	
Db	58 VQVQSVDRIILIAITE---	RTVDENDELHADSIILIYD	-----		95	
<hr/>						
Qy	109 SPAELFRGGKRTDLFQN	LRNPAPAKSRGGPGDYLLK	NYEDD----	IVRSDVALDKQKC	164	
Db	96 -----GDTEAFDE---	SPAEV--TCLDIPKLDAE	DAGGAIMRNIVALGAV--C	139		
<hr/>						
Qy	165 KIAHPDVMLELQREKAAQ	MHLVLCLKFSNTYSNLIL	TPNNVYRTHMGRAVNGSQ	LKGD	224	
Db	140 AVADFPFIENLDESLEK	-----RFGSGEQIIT---	ENNKQ-----AARLGAE	177		
<hr/>						
Qy	225 YI--OLKSLLOPIRIYS	RASLYGNIGPRKNNVIAL	DGFPMKVAGSTVDATWQH	CVIDG	282	
Db	178 YVAEFEDVTLPYELETDD	EYLLNGDEAIGMAIAGCF	FYAGYTPTTAT-----	229		
<hr/>						
Qy	283 RVKVMDFLKTRL-----	LDTLSQIRKIQKVNTYTPG	KIKWLEGVVTTTSAGGTNN	LS	336	
Db	230 --DVMEYLTRISQFGH	VVQAEDELAALNLAGAARA	GAR---SMTATSGPIDLMS	282		
<hr/>						
Qy	337 DSYAAGFLWLTMLAN	OGIDVVI	361			
Db	283 E-----TFGLVATSE	PLVI	297			
<hr/>						
RESULT	15					
C71374						
	probable glucose inhibited division protein A (gida) - syphilis spirochete					
	C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)					
	C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 26-Aug-1999					
	C:Accession: C71374					
	R:Fraser, J.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.;					
	rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.;					
	they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.					
	Science 281, 375-388, 1998					
	A>Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.					
	A:Reference number: A71250; MUID:98332770					
	A:Accession: C71374					
	A>Status: preliminary; nucleic acid sequence not shown; translation not shown					
	A:Molecule type: DNA					
	A:Residues: 1-630 <COL>					
	A:Cross-references: GB:AE001189; GB:AE000520; NID:g3322293; PIDN:AAC65038.1; PID:g332					
	A:Experimental source: strain Nichols					
	C:Genetics:					
	A:Gene: TP0044					
	C:Superfamily: gida protein					
<hr/>						
	Query Match	3.6%	Score 100.5;	DB 2;	Length 630;	
	Best Local Similarity	21.3%;	Pred. No. 12;			
	Matches	77;	Conservative	52;	Mismatches	129; Indels 103; Gaps 20;
<hr/>						
Qy	218 GSQLGK--DYIQLSKLQ	PIRIYSRASLYGNIGRPNK	NNVIALDGFPMKVAGSTVD	AVTW	275	
Db	68 GGEMGKFADACMIQ-----	YRLLNKSR--GPVAQPRIAQ	DKFL--YAQVKVYTLCTQH	118		
<hr/>						
Qy	276 QHCYLDGRVVKVMDFLK	TRLDTLSQIRKIQKVNTY	TPKKIKWLEGVVTTTSAGGT	NNL	335	
Db	119 LHLQD-----TVDVD	VCSNTTDAGYVAYGAHA	VVT--ARGRRISARAVVLTG	-----	166	
<hr/>						
Qy	336 SDSYAAGFLWL-----	NTLMGLANOGIDVVI	HRHSFFDHG-----	YNHLVDQNF	378	
Db	167 --TEMGRVYIGEYAEPE	GRGEHAEGLCAALKRKGFQ	MCRUKTGTPARVLRKSVD	---221		
<hr/>						
Qy	379 NPLFDYWLSSLYKR----	LIGPKVLAVHVAGLQRKP	RPGRVIRDKLRIYAHCTNH	NHHNY	434	
b	222 -----LSVMKEQEA	DAIMPFPFA-HV-----	EINRHADCYINTYNERHQI	263		

Qy 435 VRGSITLFIINLHRS---RKKIKLAGT-----LRDKLVHQYLLQPYGOEGLKS 479
Db 264 IRE-----NFHRSPFTSGRIKAVGTRYCPSIEDKVRKFPDRIRHQLYIEP---EGLDT 313
Qy 480 KSVQLNGQPLVMVDD-----GTLPELK-----PRDLRAGRTLVIPIPPVTMGF-FVVKNVNA 528
Db 314 EELYINGLSSCLPEDIQDEMINTIPGMERAVITRPAYAVDYAVLFPVOLGIDLQTKRVSG 373
Qy 529 L 529
Db 374 L 374

Search completed: July 30, 2002, 08:16:22
Job time: 339 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:15:13 ; Search time 36.14 seconds
(without alignments)
572.115 Million cell updates/sec

Title: US-09-836-461-2

Perfect score: 2785

Sequence: 1 MRVLCAPPEAMPSSNSRPPA.....PVTMGFFVKNVNALACRYR 534

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107	3.8	2026	1	CYAA_YEAST
2	101.5	3.6	629	1	T3MH_HAEIN
3	100.5	3.6	630	1	GIDA_TREPA
4	98.5	3.5	335	1	SWW_BOCAI
5	98	3.5	845	1	CLPC_CHLPN
6	97.5	3.5	1517	1	RPOC_CAMJE
7	96.5	3.5	969	1	DP3A_UREPA
8	96	3.4	1229	1	SLP3_YEAST
9	96	3.4	4367	1	DHHC_NEUCR
10	94.5	3.4	682	1	SNK_RAT
11	94.5	3.4	685	1	SNK_HUMAN
12	94.5	3.4	690	1	HELS_METH
13	94.5	3.4	1054	1	POL_STVMK
14	94	3.4	625	1	GIDA_LACLA
15	93.5	3.4	412	1	SYI2_BACSU
16	93.5	3.4	947	1	LKTA_PASSP
17	93	3.3	488	1	KI15_CAEEL
18	93	3.3	573	1	DPOL_MOUSE
19	93	3.3	849	1	PHSG_SYNY3
20	92.5	3.3	682	1	SNK_MOUSE
21	92.5	3.3	1616	1	VITL_CAEEL
22	92.5	3.3	2241	1	TEGU_HCMVA
23	92	3.3	461	1	GP13_YEAST
24	92	3.3	534	1	COX1_KLULA
25	92	3.3	535	1	COX1_HANWI
26	92	3.3	824	1	TC37_HUMAN
27	92	3.3	898	1	ACOC_CUCMA
28	91.5	3.3	536	1	DIT1_YEAST
29	91.5	3.3	1259	1	LN1_HUMAN
30	91	3.3	498	1	C72U_ARATH
31	91	3.3	676	1	YO7T_YEAST
32	91	3.3	1032	1	MT18_YEAST
33	90	3.2	948	1	RPOP_PODAN

ALIGNMENTS

RESULT 1

ID	CYAA_YEAST	STANDARD;	PRT; 2026 AA.
AC	P08678;		
DT	01-JAN-1988 (Rel. 06, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Adenylate cyclase (EC 4.6.1.1) (APP pyrophosphate-lyase) (Adenylate cyclase).		
DE	CYRI OR CDC35 OR HSRI OR SRA4 OR YJL005W OR J1401.		
OS	Saccharomyces cerevisiae (Baker's yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
OX	NCBI_TaxID=4932;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=86079531; PubMed=2934138;		
RA	Kataoka T., Broek D., Wigler M.;		
RT	"DNA sequence and characterization of the S. cerevisiae gene encoding adenylate cyclase.";		
RL	Cell 43:493-505(1985).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	To Van D., Perea J., Jacq C.;		
RL	Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=S288C / FY1679;		
RA	de Haan M., Smits P.H.M., Grivell L.A.;		
RL	Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.		
RN	[4]		
RP	SEQUENCE OF 1042-2026 FROM N.A.		
RX	MEDLINE=88165073; PubMed=3327602;		
RA	Masson P., Lenzen G., Jacquemin J.M., Danchin A.;		
RT	"Yeast adenylate cyclase catalytic domain is carboxy terminal.";		
RL	Curr. Genet. 10:343-352(1986).		
RN	[5]		
RP	MUTAGENESIS OF THR-1651.		
RX	MEDLINE=91122042; PubMed=1991451;		
RA	Feger G., de Vendittis E., Vitelli A., Masturzo P., Zahn R.,		
RA	Verrotti A.C., Kavounis C., Pal G.P., Fasano O.;		
RT	"Identification of regulatory residues of the yeast adenylate cyclase.";		
RL	EMBO J. 10:349-359(1991).		
CC	-!- FUNCTION: PLAYS ESSENTIAL ROLES IN REGULATION OF CELLULAR METABOLISM BY CATALYSING THE SYNTHESIS OF A SECOND MESSENGER, CAMP.		
CC	-!- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.		
CC	-!- ENZYME REGULATION: THE PRESENCE OF GTP-BOUND RAS2 PROTEIN IS REQUIRED IN ORDER TO ELICIT A MAGNESIUM-DEPENDENT ADENYLATE CYCLASE ACTIVITY.		
CC	-!- SIMILARITY: BELONGS TO ADENYLATE CYCLASE CLASS-3 FAMILY.		
CC	-!- SIMILARITY: CONTAINS 21 LEUCINE-RICH REPEATS (LRR).		
CC	-!- SIMILARITY: CONTAINS A PP2C-LIKE DOMAIN.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		

34	90	3.2	1125	1	TIE2_BOVIN	Q06807 bos taurus
35	90	3.2	1300	1	POL2_MOUSE	P11369 mus musculus
36	90	3.2	1392	1	LVS2_YEAST	P07702 saccharomyc
37	90	3.2	1406	1	CPBX_DROME	P42787 drosophila
38	89	3.2	413	1	LCAT_CHICK	P53760 gallus gall
39	89	3.2	678	1	UVRB_PASMO	P57844 pasteurella
40	89	3.2	1318	1	VP14_EBV	P03179 epstein-bar
41	88.5	3.2	286	1	PARB_CHLPN	Q927m0 chlamydia p
42	88.5	3.2	293	1	RK4_SPIOL	Q4937 spinacia ol
43	88.5	3.2	341	1	VF43_MYCTU	O10783 mycobacteri
44	88.5	3.2	859	1	ABR_HUMAN	Q12979 homo sapien
45	88.5	3.2	949	1	RCSA_ECOLI	P14376 escherichia


```

MEDLINE=98323770; PubMed=965876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete";
RL Science 281:375-388(1998).
CC -1- FUNCTION: NOT KNOWN.
CC -1- SIMILARITY: BELONGS TO THE GIDA FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001189; AAC65038.1; -
DR TIGR; TP0044; -
DR InterPro; IPR002218; GIDA.
DR Pfam; PF01134; GIDA; 1.
DR PRODom; PD003738; GIDA; 1.
DR PROSITE; PS01280; GIDA_1; 1.
DR PROSITE; PS01281; GIDA_2; 1.
DR Complete proteome.
KW
SQ SEQUENCE 630 AA; 70175 MW; 1B52C6F02C1EC275 CRC64;

Query Match 3.6%; Score 100.5; DB 1; Length 630;
Best Local Similarity 21.3%; Pred. No. 4.5;
Matches 77; Conservative 52; Mismatches 129; Indels 103; Gaps 20;

Qy 218 GSQLGK--DYIQLKSLQPIRIYSRASLYGNIGPRKNVIALLDGFMKVAGSTVDVATW 275
|:::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 68 GGEMGFADACMIQ-----YRLNKSRR--GPAVQAPRIQADKFL--YAKVKYTLLECTQH 118
|:::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Qy 276 QHCYIDGRVVKVMPDLKTRLLDTSQIRKIQKVVNTVTPGKKIWLEGVVTTSAGGTNNL 335
|:::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 119 LHLYQD---TVVDVVCSTNTDAGVYVAAHAVVT--ARGRRISARAVLTTG----- 166
|:::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Qy 336 SDSYAAGFLWL-----NTLGLMANQIDVIVIRHSFFDHG-----YNHLVDQNF 378
|:::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 167 --TFMEGRVYIGEYEAPEGRLGEHAAGLAKRKKGFGOMGLKGTGPARVLRKSVS--- 221
|:::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Qy 379 NPLPDYWLSLLYKR---LIGPKVLAVHVAGLQKRPGRVIRDKLRIYAHCTNNHHNY 434
|:::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 222 -----LSVMEKQADAIIMRPFSA-HV-----EINRPADCYINTNERTHQL 263
|:::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Qy 435 VRGSTTLFIINLHRS---RKIKLAGT-----LRDKLVHQYLLQPYGOGLKS 479
|:::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 264 IRE-----NFRHSPPFSGRIKAGTVGTCPSIEDKVRKFPDRIRHQLYIEP---EGLDT 313
|:::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Qy 480 KSVQLNGQLVMVDD-----GTLPCLK----PRPLRAGRTLVIPIPTMGP-FVVKNVNA 528
|:::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 314 EELYINGLSCLPEDIQDEMINTIPGMRVAVITRPAYADVAVLPVOLGIDLQTKRVSG 373
|:::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Qy 529 L 529
Db 374 L 374

RESULT 4
SYW_BUCAI
AC SYW_BUCAI STANDARD; PRT; 335 AA.
ID PS7602;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE DE Tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Tryptophan--tRNA ligase)
DE (TrpRS).
GN TRPS OR BU536.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TOKYO 1998;
RX Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RA "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -!- CATALYTIC ACTIVITY: ATP + L-tryptophan + tRNA(Trp) = AMP +
CC diphosphate + L-tryptophanyl-tRNA(Trp).
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP001119; BABI3229.1; -.
DR InterPro; IPR001412; trna-synt_I.
DR InterPro; IPR002306; trna-synt_trp.
DR Pfam; PF00579; trna-synt_1b; 1.
DR PRINTS; PR01039; TRNASYNHTRP.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
DR Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 14 22 "HIGH" REGION.
FT SITE 196 200 "KMSKS" REGION.
FT BINDING 199 199 ATP (BY SIMILARITY).
FT SEQUENCE 335 AA; 38549 MW; 871562D5A6734E3F CRC64;
QY 151 IYRSDVALDKQCKIAQHPDVMLELQREKAAQMHVLVLEKEQFSNTYSNLIITPEPNYRT 210
DB 132 LMAADILLYQTNFVPVQDQKHVELTRN-----IAHFNLSLYGH-VETLPKPLIT 181
QY 211 MHGRAVNSQLGKDYIQLKSLQPIRIYSRASLYGNIGRPKNVYALDGG---FMKVA 266
DB 182 QHGS-----KIMSLEPSKMSKSDI-----NKKNVIFLLDDDKTIVISKIQ 222
QY 267 GSTVDVAVTQHCVID-----GRVVKVM-----DPL-----KTRLLDPL 299
DB 223 NAYTSETSKYIYDEKPKGINSNLEILSALTNRKDIDILLKELEGMYSEFKNIVADHL 282
QY 300 SDQIRKIQKVNTYTPG-----KKIWLEGVVTTSAGTNNLSDSYAAGFWLNTLGLM 352
DB 283 SKFLYKQLQSYNDRNDEVYLLAKIAGVAGAMKSLKSNKLTTRKY-----DKLGLI 332
RESULT 5
CLPC_CHLPN
ID CLPC_CHLPN STANDARD; PRT; 845 AA.
AC Q9Z8A6; Q9K297; Q9JSE9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable ATP-dependent Clp protease ATP-binding subunit.
GN CLPC OR CPN0437 OR CP0316.

```

```

OS Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S., trachomatis.";
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heldelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linhorst K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -!- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY. CLPC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001626; AADI8581.1; -.
DR EMBL; AE002194; AAF38172.1; -.
DR EMBL; AP002546; BAA98645.1; -.
DR PHCI-2DPAGE; Q9Z8A6; -.
DR TIGR; CP0316; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR003959; AAA_subfam.
DR InterPro; IPR001270; CLP_AB.
DR InterPro; IPR004176; Clp_N.
DR InterPro; IPR001943; UVR.
DR Pfam; PF00004; AAA; 1.
DR Pfam; PF02861; Clp_N; 2.
DR Pfam; PF02151; UVR; 1.
DR PRINTS; PR00300; CLPPROTEASEA.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00570; CLPAB_1; 1.
KW Chaperone; ATP-binding; Repeat; Complete proteome.
FT NP_BIND 232 239 ATP (POTENTIAL).
FT NP_BIND 569 576 ATP (POTENTIAL).
FT CONFLICT 4 4 K -> T (IN REF. 3).
FT CONFLICT 141 141 K -> R (IN REF. 1).
FT SEQUENCE 845 AA; 94980 MW; C9D9205458812866 CRC64;
QY 13 SSNSRPPACLPAGALYALLHLHLSSQAGDR-----RPLPVD-----RAA 53
DB 160 SSRSNPFSSKSP-----LGHSLGSDKNEKLSALKAYGYDLTEMVRESKLDPPVIGRSS 211
Query Match 3.5%; Score 98; DB 1; Length 845;
Best Local Similarity 19.8%; Pred. No. 11;
Matches 131; Conservative 89; Mismatches 241; Indels 202; Gaps 30;

```

```
QY 54 GLKEKTLILLDVSTKNPV-----RTVNFNLSQLDPSIIHDGWLDFLSKRLVTLARGLS 109
Db 212 EVERILILCLRRRKNPNVLIGAGVGKTAIVEGLAQKIIILAEVDPALRRKRLITLIDLALM 271
QY 110 PAFIRFGG-----KRTDFLOF-----ONLRNPAKSR- 135
Db 272 IAGTKYRQOFERIKAVMDYVRKHCNILLFIDELHTIVGAGAAEADASNILKPALARG 331
QY 136 -----GGPGDYUKNYEDDIRSDVALDKQCKIAQHP-----DVMELQR-----EKAQCM 184
Db 332 ETQICGATTIDEYRKHIE-----KDAALER-RFQIVVHPSPVDETIELRGLKKKYEHEH 385
QY 185 HLVLKQFQSNYSNLIITENPNYTMGR-----AVNGSQLKQDIQ 227
Db 386 HNVFITEEALKAAATL-----SDQYVHGRFLPKDAIDLLDEAGARVRVNTMGQPTDLMK 439
QY 228 LKSLQPIRIYRSASLYGPNIGRPRKNVIALLDGPMKVGSTVDVAV--TWQH-----CY 279
Db 440 LEAEIENTKLAKEA-----IGTQYEKAAGLRDEKKLRERLQSMKQEWENHKEHQVP 494
QY 280 IDGRVYKVMDFLKT-----RLDLTSDQIRKIQKVNTYTPGKKIWLLEGVVTTSAGGTNN 334
Db 495 VDEEAVAQWVSLQTGIPARSARTEASEKLLKLEDTLRRKVIQNDVAVTSICRAIRSRRTG 554
QY 335 LSDSY--AAGFLWLNTLG-----MLANQ-----GIDVIRHSFFDH----- 368
Db 555 IKDPNRPFGSFLGPTGVGKSLAQAIEMFGGEDALIOVDMSEYMEKFAATRMGMS 614
QY 369 -GY-----NHLVQD-----NFNP-LPDYWLSLYK-RL---IGPKVLAV 401
Db 615 PGYVHEGEGHLEQVRRRCVVLDFEIEKAHPDMDLMQLQILEQGLTDSFGRKVD 674
QY 402 HVAGLQRKPRGVRIRDKLRIYAHCTNHHNHNHVRGSIITLFIINLHRSRKKIKLAGTLRD 461
Db 675 HAILIMTSLNGLADLRKSGEIGFGLKSHMDYKVIQEKIE-----HAMKKHLK--PEFIN 726
QY 462 KLVHYLLOPQGEQ-----LKSQSVQNLGQPLV---MVDDGTLPPELKR 503
Db 727 RLDESIVPRPLEKESLSBIIHLEINKLSRLKNYQMALNPDSVIFLVTGKHSPEMGAR 786
QY 504 PLR 506
Db 787 PLR 789
RESULT 6
RPOC_CAMJE STANDARD; PRT; 1517 AA.
AC Q9PI30;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase
beta' chain) (RNA polymerase beta' subunit).
GN RPOC OR C00479.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltham T., Holroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrett B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
reveals hypervariable sequences."
RL Nature 403:665-668(2000).
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
```

```
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -!- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
CC BETA' CHAIN.
CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL139075; CAB75117.1; -
CC InterPro; IPR000722; RNA_pol_A.
CC Pfam; PF00623; RNA_pol_A; 1.
CC Transfaser; DNA-directed RNA polymerase; Transcription;
CC Complete proteome.
CC
CC SQ SEQUENCE 1517 AA; 168821 MW; 6778D27483AB70EC CRC64;
Query Match 3.5%; Score 97.5; DB 1; Length 1517;
Best Local Similarity 20.3%; Pred. No. 26;
Matches 111; Conservative 87; Mismatches 188; Indels 161; Gaps 31;
QY 46 PLPVDRAAGLKEKTEKTLILLDVSTKNPVRTVNFNLSQLDPSIIHDGWLDFLSKRLVTLA 105
Db 442 PVMNARA-----PTLHKLSIQAFHEPVLVEGK---AQIHLPLCAAFNDFDGDQMAVHP 493
QY 106 RGLSPAFLRFGGKRTDFLOFQNLNPA--KSRGGPGDYUKNYEDDIRSDVALDK--Q 161
Db 494 --LSEAI--ACEKVLMLSSMILLPASCKSVTVPSQDMVLGIY-----LSLEKAGA 542
QY 162 KGC-KIAQHPDYMLELOREKAAQMHLLVLLKEQFSNTYSNLIITPENNYRTHMGRVNGSQ 220
Db 543 KGSKHCITGIDVEMMALESKCLDIHA-----SIQTMVDGKRKITTAGRLIVKSI 591
QY 221 LKQDYIQLKSLLOPIRIYRSASLYGPNIGRPRKNVIALLD-----GPMKVGSTVDVAVT- 274
Db 592 L-PDFVPENSNWNVKLVK-----KKDIAALVDYVYKQGGLEITASFGLDLKN 635
QY 275 --WQHCYIDGRVVKVMDFL---KTRLLDLSQIRKIQKVNTYTPGKKIWLLEGVVTTS 328
Db 636 LGFEYATKAGISISTADIIVPNDKQKAIDEAKKQVREIQ---NSYMLG-----LITS 684
QY 329 AGGTNNLSDSYAAGFLWLNTLGLANQGDIVVIRHSFFDGHYNHLVDQNFNPLPDYWL 388
Db 685 GERYNKIID-----IKSTNNVLSKEMKLVKDK--EGFN-----SI 720
QY 389 LYKRLIGPKVLAVHVA-----GLQKKPRPGRVIRDKLRIYAHCTNHHNHNHVRG-SITL 441
Db 721 YMMADSGARGSAQISQLAAMRGLMTKP-DGSIIITPI-----ISNREGNLVLE 769
QY 442 FIINLHRSRK-----KIKLAGTLRDKLIVH---QYLLQPYQOGLSKSVQNLGQPL 489
Db 770 YFISTHGARKGLADTALKTANAGYLTRKLIDVAQNVKITIEDCGTH-----EGVEINE--- 822
QY 490 VMVDDGTLPPELPRPLRAGRTL---VIPPVT-----MGFFVVKNVN- 527
Db 823 ITADSSIIITLEERIL--GRVLAEDVIDPITNSVLVFAEGLTMDKEKAKILGESIGKSVNI 880
QY 528 --ALACR 532
Db 881 RTPITCK 887
RESULT 7
DP3A_UREPA STANDARD; PRT; 969 AA.
ID DP3A_UREPA
```

Q9PQ74;
16-OCT-2001 (Rel. 40, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
DNA polymerase III alpha subunit (EC 2.7.7.7).
DNAE OR U4415.
Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
Mycoplasmataceae; Ureaplasma.
NCBI_TaxID=134821;
[1]
SEQUENCE FROM N.A.
RC STRAIN=SEKOVAR 3; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Cassell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
urealyticum".
RL Nature 407:757-762(2000).
CC -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC THE ALPHA CHAIN IS THE DNA POLYMERASE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate -> N diphosphate
+ [DNA](N).
CC -!- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA
CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE
DIMERIZATION TO FORM THE POLIII' COMPLEX. POLIII' ASSOCIATES WITH
THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, DELTA', PSI,
AND CHI) AND WITH THE BETA CHAIN (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-C FAMILY. DNAE
SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AB002138; AAF30826.1; -
CC InterPro: IPR003141; PHP.N.1
CC Pfam: PF02231; PHP.N.1
CC SMART; SM00481; POLIIIAC.1
CC Transferase: DNA-directed DNA polymerase; DNA replication;
CC Complete proteome.
CC SEQUENCE 969 AA; 112343 MW; 0569486460E2960B CRC64;
Query Match 3.5%; Score 96.5; DB 1; Length 969;
Best Local Similarity 20.4%; Pred. No. 17;
Matches 112; Conservative 67; Mismatches 172; Indels 199; Gaps 29;
QY 47 LPVDRAAGLKEKTELI---LIDVSTK--NPVRTVNFNLSLQDPSIHDGWLDFLSSKRL 101
DB LAIKKNTLKEWYLLHKELEFISKLINAPRQIGTHAAGILSDLI-----SNI 453
QY 102 VTLARGLSPAFL-----RFGGKRTDFLQFNL-----RNPASRGGPGPD 141
DB 454 IPTQLGINDRPLSQYSNEYLEREGLKMDLLGLKNTIIDNVLMKTYENQNK-----ID 508
QY 142 YLKNVEDDIDVRSDDVALDKQCKIAQHPDVMLELQRAQMHVLLKAEQFSNTYSNLI 201
DB 509 LFNINYNDFEEDLAKARTNGIFQLESEG-----MKVLLK----- 545
QY 202 LTPBNRYRTMHGRAVNSQLGKDYIQLKSLLOPIRTYRSALYGPNGRPNKVIALL-- 259
DB 546 -VRPQN-----IEDISIVS--ALFRPG---PQONIKTFVER 575
QY 260 -----DGFMKVAGSTVDATVWQHCYIDGRVVKVMDFLKT--RLDLTSLSDIRK- 305

Db 576 RFKREPSYWNATRKILEPTYGIIVQE-----QVIELVKVTIANEDIATSDNFERRA 627
QY 306 IOKVNTYTPGKKTWLEGVVTTSAGTNNLSDSYAAGFLMINTGLMLANQIDVIRHSF 365
DB 628 ISK-----KDEKILIQ-----LKDDFINGALNNNYKOPLVNOIFYEYF--SF 667
QY 366 FDHCYNHLVDQNFNPLPDYWLSSL--YKRLIGPKVLAVHVAGLQKRPGRVIRDKLRIY 423
DB 668 AHYGNFNSHSLAYSII--SYWLAYLKHYYTLEFLSVLLSHTS-----ASKDKLLSY 716
QY 424 AHCTNHHNNHYVRSITL-----FIINLHRS-----RKKIKLA---GTLR 460
DB 717 LNEAKEFNISIKGPDIOQVFSNDFIDTQKQIIRFEKTIKFGDELLKKIKSALQKSTLS 776
QY 461 DKLHVQYLLQYQGEGLKSKSVQL--NGQPLVMVDDGTLPELKPRPLRAGRITLVPPVTMG 519
DB 777 DYI--SYI-----DALKKGNVSLKNIEILRV--GTDFDSFGINRL----- 812
QY 520 FFVVKVNNAL 529
DB 813 -FLNLLNEI 821
RESULT 8
SIP3_YEAST STANDARD; PRT; 1229 AA.
ID SIP3_YEAST
AC P38717;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SIP3 protein.
GN SIP3 OR YNL257C OR N0844.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
ON NCBI_TaxID=4932;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=94173726; PubMed=8127709;
RA Lesage P., Yang X., Carlson M.;
RT "Analysis of the Sip3 protein identified in a two-hybrid screen for
interaction with the SNF1 protein kinase.";
RL Nucleic Acids Res. 22:597-603(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FV1679;
RX MEDLINE=97377992; PubMed=9234673;
RA Sen-Gupta M., Gueldener U., Beinhauer J.D., Fiedler T.A.,
RA Hegemann J.H.;
RT "Sequence analysis of the 33 kb long region between ORC5 and SU11
from the left arm of chromosome XIV from Saccharomyces cerevisiae.";
RL Yeast 13:849-860(1997).
CC -!- FUNCTION: INTERACTS WITH THE SNF1 PROTEIN KINASE.
CC -!- SIMILARITY: STRONG, TO YEAST YHR155W.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; U03376; AAA17885.1; -
CC EMBL; X96722; CAA65487.1; -
CC EMBL; Z71533; CAA96164.1; -
CC PIR; S42391; S42391.
CC SGD; S0005201; SIP3.
CC InterPro: IPR001849; PH.
CC Pfam: PF00169; PH.1.
CC SMART; SM00233; PH; 1.

DR	InterPro: IPR004273; Dynein_heavy.
DR	InterPro: IPRO01482; GSPIL_E.
DR	Pfam: PF03028; Dynein_heavy.1.
DR	Pfam: PF00437; GSPIL_E.1.
KW	Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.
FT	DOMAIN 676 693 COILED COIL (POTENTIAL).
FT	DOMAIN 1176 1215 COILED COIL (POTENTIAL).
FT	DOMAIN 1327 1351 COILED COIL (POTENTIAL).
FT	DOMAIN 1557 1574 COILED COIL (POTENTIAL).
FT	DOMAIN 1637 1668 COILED COIL (POTENTIAL).
FT	DOMAIN 2045 2073 MICROTUBULE-BINDING (POTENTIAL).
FT	DOMAIN 2195 2218 COILED COIL (POTENTIAL).
FT	DOMAIN 3193 3296 COILED COIL (POTENTIAL).
FT	DOMAIN 3423 3481 COILED COIL (POTENTIAL).
FT	DOMAIN 3778 3809 COILED COIL (POTENTIAL).
FT	NP_BIND 1943 1950 ATP (POTENTIAL).
FT	NP_BIND 2240 2247 ATP (POTENTIAL).
FT	NP_BIND 2605 2612 ATP (POTENTIAL).
FT	NP_BIND 2947 2954 ATP (POTENTIAL).
SQ	SEQUENCE 4367 AA; 495568 MW; 1E10F3E2D170D6DF CRC64;

Query Match		3.4%; Score 96; DB 1; Length 4367;
Best Local Similarity		21.1%; Pred. No. 1.5e+02;
Matches 110; Conservative 57; Mismatches 179; Indels 176; Gaps		

QY	33	LHLSLSQAGRRPL-----PVDRAAGLKEKTLILDVSTKNPVTVNENFLSLQLDP	SI	87
DB	3548	LHLS-GVFQKHNPWEYLSTADERSWGENTLPVDDLCETENAIIKKFRNYPLIIDPSG		3606
QY	88	IHDGWLDLSSKRVLARGLSPAF-----LRFGGKRTDFLOFNLRPNAKSRGGPGP		140
DB	3607	RATEEFNLRESKDKLIIVTSFLDSDSFVKLESSLRG-----NPILIODAEHL		3653
QY	141	DYYL-----KNYEDDIVRSDVALDGOKGCKIAHQHPDVMLEIQ-REKAQMHLVLLKEQFS		194
DB	3654	DPVLNHVNLKEYQTGGRVLIQLGKO---QIDFSPAFLYLSTRDPSATFAPDICSR---		3707
QY	195	NTYSNLTITPENNYRTMHGRAVNSQLGKDYLQLKSLLQPIRIYSRASLYGNIGRPKN		254
DB	3708	TTFVNFTVTQSS-----LQTSLENEVLKSER-----PDVDERSN		3742
QY	255	VI-----ALIDGFMKVAGSTVDAVTWQHICYIDGRVVKVMDLFKLRLDLTL		299
DB	3743	LIKLOGEFKVHLRQLEKLLQALNESRGNIID-----DDHVLETLTKEAAEIS		3793
QY	300	SDQIRKIQKVNTYTGGKIKIWLEGVTTTSAGGT---NNLSDSYAAGFWLWTLGLMANQG		356
DB	3794	A-----KMSNT-----EGVMAEQEITLQYNIIARSCSAVFVLEQL-----		3830
QY	357	IDVVIRHSFPDHGYNHLVDQNFNPLPDYWLSSLYKRLGPKVLAVHVAGLORKPRGRVI		416
DB	3831	-----HYLNHF-----YRFSQY-----FLDTFHSVLRGNP-----		3856
QY	417	RDKLRIYAHCTNHNHNHYVRGSIT--LFTINLHRSRKKIKLAGTLURDKLVHGYLL---		471
DB	3857	-----HLANETHNVREDIIVKDLFYATF----KRTALGLLKQDRITLAMLLAQASP		3904
QY	472	YGQE-GL-----KSKSVOLNGQPVLVMWDGTLPELK		501
DB	3905	YKMDKGLLDIILDERIEGDKVSIDQNTREEAFARAKKIPALK		3946

RESULT	10
SNK_RAT	SNK_RAT
ID	SNK_RAT
AC	STANDARD; PRT; 682 AA.
DC	Q9R012;
DT	16-OCT-2001 (Rel. 40, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)
DE	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Serine/threonine-protein kinase SNK (EC 2.7.1.-) (Serum inducible kinase).
DE	DNK

OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9452760; PubMed=10523297;
 RA Kauselmann G., Weiler M., Wulff P., Jessberger S., Konietzko U.,
 RA Scalfidi J., Staubli U., Bereiter-Hahn J., Streibhardt K., Kuhl D.;
 RT "The polo-like protein kinases Fnk and Snk associate with a Ca(2+)-and
 RT integrin-binding protein and are regulated dynamically with synaptic
 RT plasticity.";
 RT EMBO J. 18:5528-5539(1999).
 RL -!- FUNCTION: MAY PLAY A ROLE IN THE DIVISION OF AT LEAST SOME CELL
 CC TYPES, SUCH AS FIBROBLASTS, AND COULD FUNCTION IN EMBRYOGENESIS,
 CC WOUND HEALING OR NEOPLASIA (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC CDC5/POLO SUBFAMILY.
 CC -!- SIMILARITY: CONTAINS 2 POLO BOX DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF136583; AAF08366.1; -.
 DR HSP; P00518; 1PKH.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR000959; POLO_box.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR SMART; SM00220; S_TKc; 1.
 DR PROSITE; PS50078; POLO_BOX; 2.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR Transferase; Serine/threonine-protein kinase; ATP-binding; Repeat.
 KW DOMAIN 54 59 POLY-HIS.
 FT DOMAIN 79 331 PROTEIN KINASE.
 FT NP_BIND 85 93 ATP (BY SIMILARITY).
 FT BINDING 108 108 ATP (BY SIMILARITY).
 FT ACT_SITE 202 202 BY SIMILARITY.
 FT DOMAIN 507 570 POLO BOX 1.
 FT DOMAIN 603 674 POLO BOX 2.
 FT SEQUENCE 682 AA; 77919 MW; 58C50DEBDE83D5F3 CRC64;
 CC
 CC Query Match 3.4%; Score 94.5; DB 1; Length 682;
 CC Best Local Similarity 22.6%; Pred. No. 15;
 CC Matches 95; Conservative 58; Mismatches 166; Indels 101; Gaps 21;
 QY 129 RNPASRGPGPDYLYKNYEDDIVRSVDALDKQCKGKIAQHPDVMLELQREAAQ-----183
 DB 81 RGVVLGGG-----FACYE-----WDLTNKKVYAAKLIHPSRVAPHQREKIDKELEH 131
 QY 184 -----MHLVLKQFSNTYSLNLTPEPNNTYRTMHGRAVSGQKGYIQLKSLQIP-IRI 237
 DB 132 RILHKKHVQFHYFEDKENIYILLEYCSRRSM-----AHLK---ARKVLTEPEVRY 181
 QY 238 YSRASLYGPNIGRPNKVNIALLDGFMKAVAGSTVDVNTWQHCHYIDGRVVKVMDFLKTRLLD 297
 DB 182 YLRQIVSGLKYLHEQE-----ILHRDLKLGNNFFN-----EAMELKVGDFGLAARLE 228
 QY 298 TLDQIRKIKQVNTYTP-----GKKIWLEGVV---TTSAG-----GTNNLSDSY--- 339
 DB 229 PLEHRRRTICGTPVLSPEVLNKGHCESDIWALGCMVTMLIGRPPFTTNLKETYRC 288
 QY 340 --AGFLWLNTL-----GMLANQ-----GIDVVIRHSFDPHG--NHLVDQNFNPL 361
 DB 340 --AGFLWLNTL-----GMLANQ-----GIDVVIRHSFDPHG--NHLVDQNFNPL 361

DB 289 IREARYTMPSSLAPAKHLIASMLSKNPEDRPSLDDIIRHDFFLQGTTPDRLSSSCCHTV 348
 QY 382 PDYWLSLLYRLGPKVLAVHVGLOKRP-----PGRVIRDKLRIYAHCTNHNNHYVRG 437
 DB 349 PDFHLSSPAKNFP-KKAAALFGGKKDKARYNDTHNKVSKEDDIY-----KLKRDHLKKT 402
 QY 438 SITLFIINLHRSRKKIKLACTLRDLKLVHQLQPYGOEGLKSKSVQNGOPLVWDDGTL 497
 DB 403 SITQ-QFSKRTDEELQPPPTFAK-----SGTSAVENK--QQIGDAIRMIVRGTL 450
 RESULT 11
 SNK_HUMAN
 ID SNK_HUMAN STANDARD; PRT; 685 AA.
 AC Q9NYX3; O60679; Q9UE61;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Serine/threonine-protein kinase SNK (EC 2.7.1.-) (Serum inducible
 DE kinase).
 GN SNK.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ouyang B., Dai W.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon;
 RA Anderson K.M., Nerurkar S.S., Hansbury M.J., Fornwald J., Scott G.,
 RA Bouzyk M., Mui P., Imbruglia C.S., Carlson K., Marshall L.A.,
 RA Roshak A.K.;
 RT "Identification and characterization of human serum-inducible kinase
 RT (SNK), a novel member of the polo-kinase family of cell cycle
 RT regulators: potential implication for regulation of vascular smooth
 RT muscle proliferation.";
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 110-408 FROM N.A.
 RA Fidler C., Boulwood J., Wang Jabs E., Wainscoat J.S.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: MAY PLAY A ROLE IN THE DIVISION OF AT LEAST SOME CELL
 CC TYPES, SUCH AS FIBROBLASTS, AND COULD FUNCTION IN EMBRYOGENESIS,
 CC WOUND HEALING OR NEOPLASIA (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC CDC5/POLO SUBFAMILY.
 CC -!- SIMILARITY: CONTAINS 2 POLO BOX DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF059617; AAC14573.1; -.
 DR EMBL; AF223574; AAF62897.1; -.
 DR EMBL; U85755; AAD00375.1; -.
 DR HSP; P00518; 1PKH.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR000959; POLO_box.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00659; POLO_box; 2.
 DR SMART; SM00220; S_TKc; 1.
 DR PROSITE; PS50078; POLO_BOX; 2.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.

Transferase; Serine/threonine-protein kinase; ATP-binding; Repeat.
KW DOMAIN 57 64 POLY-HIS.
FT DOMAIN 82 334 PROTEIN KINASE.
FT NP_BIND 88 96 ATP (BY SIMILARITY).
FT BINDING 111 111 ATP (BY SIMILARITY).
FT ACT_SITE 205 205 BY SIMILARITY.
FT DOMAIN 510 573 POLO BOX 1.
FT DOMAIN 606 677 POLO BOX 2.
FT CONFLICT 28 28 G -> A (IN REF. 2).
SQ SEQUENCE 685 AA; 78222 MW; DECBE929CA35A412 CRC64;

Query Match 3.4%; Score 94.5; DB 1; Length 685;
Best Local Similarity 23.1%; Pred. No. 15;
Matches 84; Conservative 46; Mismatches 144; Indels 89; Gaps 18;

QY 129 RNPASRGPGPDYLLKNEYDIDVRSVALDKCKGCKIAQHPDVMLQREKAAQ-----183
DB 84 RGVLGKGG-----FAKCYE-----WTDLTNNKYVYAAKLIIPHSRVAKPHQREKIDKEIELH 134
QY 184 -----MHLVLKKEQFSNTYSNLIITPEPNRYTMHGRAVNGSQLGKDYIQLKSLQP-IRI 237
DB 135 RILHHKHVVQVHYVEDKENIYVILEYCSRSM-----AHILK---ARKVLTEPEVRY 184
QY 238 YSRASLYGPNIGRPRKNVIALDGMKAVAGSTVDVATWQHCHYIDGRVVKVMDFLKRLLD 297
DB 185 YLRQVSLGKYLHQE-----ILHRDLKLGNNFIN-----EAMELKVGDFGLAARLE 231
QY 298 TISQIRKIQKVNTYTP-----GKKIWLGVV--TTSAG-----GTNNLSDSY-- 339
DB 232 PLEHRRVTCGPNVLSPEVLNKGQHGCESDIHALGCWYMTLGRPPFTTNLKETRC 291
QY 340 --AAGFLWNL-----GMLANQ-----GIDVIRHSHFFDHGY--NHLVDQNFNPL 381
DB 292 IREARYTMPSSLLAPAKHLLIASMLSKPNEDRPSLDIIRHDFLQGTPTDRLSSCCCHTV 351
QY 382 PDYWLSLYKRLGPKVLAVHAGLQRPKPR---PGRVIRDKLRIYACTNNHNNYVRG 437
DB 352 PDFHLSLSPAKNFF-KKAAALFGGKKDKARYIDTHNRVSKEDEDIY-----KLRHDLKKT 405
QY 438 SIT 440
DB 406 SIT 408

RESULT 12
HEL5_METTH STANDARD; PRT; 690 AA.
AC 026901;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative ski2-type helicase (EC 3.6.1.-).
GN MTH810.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanothermobacter.
OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RA MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -I- SIMILARITY: BELONGS TO THE HELICASE FAMILY. SKI2 SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AE000858; AAB85310.1; -
DR InterPro; IPR001410; DEAD
DR InterPro; IPR003583; HHH1.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR SMART; SM00278; HHH1; 1.
KW Hypothetical protein; Hydrolase; Helicase; ATP-binding;
KW Complete proteome.
FT NP_BIND 45 52 ATP (POTENTIAL).
FT SITE 133 136 DEXH BOX.
FT SITE 133 136
SQ SEQUENCE 690 AA; 78509 MW; 90BEE6110C4133F3 CRC64;

Query Match 3.4%; Score 94.5; DB 1; Length 690;
Best Local Similarity 21.8%; Pred. No. 15;
Matches 79; Conservative 52; Mismatches 112; Indels 119; Gaps 18;

QY 22 LAPGALYLALLHLHLS-LSSQAG-----DRRLPLVDRAA-----GLKEKTLILLDV 65
DB 157 LNPVSRIVALSATLSNMDEIAGWLDARVVEHDYRPVPLHREVLDTEMFGVREKNDVWLK 216
QY 66 STKNPVRTVNENFLSLQIDPSIIHDG--WLDFLSKRLV-TLARGLSPAFLRFGKRTDF 122
DB 217 LERS-----LEDGSQLAFVSTRTTESLASHAD---KISKIPDD 255
QY 123 LOFQNLNPA-----KSRGGPGDYLLKNEYDDIVRSVALDKCKGCKIAQHPDVMLE 175
DB 256 M-VESFREAVAGKLVKPSRSGSPPTSTCLK-----LAECLEAGIAFHAGLFN 302
QY 176 LOREKAAQMHVLKKEQFSNTYSNLIITPEPNRYTMHGRAVNGSQLGKDYIQLKSLQPI 235
DB 303 RQRE-----IIEDFRQGNILMITATPS---LMYG-----VNLPSRTVVI 339
QY 236 RIYSRASLYGP-----NIGRPRKNVIALDGF-MKVAGSTVDVATWQHCHYID 281
DB 340 RDTWTWTSQPRIPVFDYEQMSGRAGRPQYDDA---GYSYLIARSHDEAMDLEYYIR 395
QY 282 GRVVKVMDFLKTRLLDTLSQIRKIOKVNVNTYTPGKKIWLEGVVTTSAGTNNLSDSYAA 341
DB 396 GEVERT---TSRIETNRDALYRQI-----IAQVASGLSGTTEELADFFRN 437
QY 342 GF 343
DB 438 TF 439

RESULT 13
POL_SIVMK
ID POL_SIVMK STANDARD; PRT; 1054 AA.
AC P05697;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE POL polyprotein (Contains: Protease (Retropepsin) (EC 3.4.23.-);
DE Reverse transcriptase (EC 2.7.7.49); Ribonuclease H (EC 3.1.26.4)).
GN POL.
OS Simian immunodeficiency virus (K6W isolate) (SIV-MAC).
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11735;
RN [1]
RP SEQUENCE FROM N.A.


```

RESULT 15
SYTZ_BACSU          STANDARD;          PRT;    412 AA.
ID SYTZ2_BACSU
AC P25151;
CT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-OCT-2001 (Rel. 40, Last annotation update)
DE Tyrosyl-tRNA synthetase 2 (EC 6.1.1.1) (Tyrosine-tRNA ligase)
DE (TYRS 2).
DE TYRZ OR TYRR OR TYRS1 OR TYRT OR IPA-9R.
OS Bacillus subtilis.
OS Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=922116127; PubMed=1806041;
RA Glaser P., Kunst F., Debarbouille M., Vertes A., Danchin A.,
RA Dedonder R.;
RT "A gene encoding a tyrosine tRNA synthetase is located near sacS in
RT Bacillus subtilis.";
RL DNA Seq. 1:251-261(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7934828;
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Mosser I.,
RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees.";
RL Mol. Microbiol. 10:371-384(1993).
CC -1- CATALYTIC ACTIVITY: ATP + L-tyrosine + tRNA(Tyr) = AMP +
CC diphosphate + L-tyrosyl-tRNA(Tyr).
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collabora-
CC between the Swiss Institute of Bioinformatics and the EMBL outstat-
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for commo-
CC entities requires a license agreement (See http://www.isb-sib.ch/anno-
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X52480; CAA36724.1;
DR EMBL; X73124; CAA51565.1; ALT_INIT.
DR EMBL; Z99123; CAB15872.1; ALT_INIT.
DR PIR; S16426; S16426.
DR PIR; S39664; S39664.
DR HSSP; P00952; 4TS1.
DR SubtLList; BG10555; tyrZ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR002305; tRNA-synt_1b.
DR InterPro; IPR001412; tRNA-synt_I.
DR InterPro; IPR002307; tRNA-synt tyr.

```

Result No.	Score	Query %			DB	ID	Description
		Match	Length				
1	2785	100.0	534	4	Q9HB38	homo sapien	
2	2746	98.6	592	4	Q9HB37	homo sapien	
3	2478	89.0	480	4	Q9HB39	homo sapien	
4	953	34.2	535	11	Q9QZF8		
5	937	33.6	545	4	Q9UL39		
6	932.5	33.5	543	4	Q9I251		
7	932	33.5	545	6	Q9MYX0		
8	836.5	30.0	523	13	Q90YK5		
9	302	10.8	521	10	Q9SDA1		
10	302	10.8	543	10	Q9FF10		
11	261.5	9.4	516	10	Q9FLK8		
12	256.5	9.2	527	10	Q9LRC8		
13	246	8.8	536	10	Q9FZP1		
14	171	6.1	935	5	Q9VE79		
15	155	5.6	190	10	O82604		
16	113	4.1	4533	5	Q9B1X3		

QY 181 AAQHLVLLKEQFSNTYSNLILTEPNNTYTMHGRAVNSQLGKDYIQLKSLLOPIRIYSR 240
DB 181 AAQHLVLLKEQFSNTYSNLILTEPNNTYTMHGRAVNSQLGKDYIQLKSLLOPIRIYSR 240
QY 241 ASLYGPNIGRPRKNNVIALLDGFMKVAGSTVDVATWQHICYIDGRVVKVMDFLKTRLLDLS 300
DB 241 ASLYGPNIGRPRKNNVIALLDGFMKVAGSTVDVATWQHICYIDGRVVKVMDFLKTRLLDLS 300
QY 301 DOIRKIQVNTYTPGKKIWLEGVVTTSAGGTNNLSDSYAAGFLWNLTLGLMANGIDV 360
DB 301 DOIRKIQVNTYTPGKKIWLEGVVTTSAGGTNNLSDSYAAGFLWNLTLGLMANGIDV 360
QY 361 IRHSFDDHGYNHLVDQNFNPLDYWLSLLYKRLIGPKVLAVHVGAGLQKRPGRVIRDKL 420
DB 361 IRHSFDDHGYNHLVDQNFNPLDYWLSLLYKRLIGPKVLAVHVGAGLQKRPGRVIRDKL 420
QY 421 RIYACTNHHNNHNVYRGSTITLFIINLHRSRKKIKLAGTLRDKLVHQLVYLLQPYGQGLKSK 480
DB 421 RIYACTNHHNNHNVYRGSTITLFIINLHRSRKKIKLAGTLRDKLVHQLVYLLQPYGQGLKSK 480
QY 481 SVQLNGQPLVMVDDGTLPELKPRLRAGRTLVIPTVMGFFVKNVNALACRYR 534
DB 481 SVQLNGQPLVMVDDGTLPELKPRLRAGRTLVIPTVMGFFVKNVNALACRYR 534

RESULT 2
QY 09HB37 PRELIMINARY; PRT; 592 AA.
AC 09HB37;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE HEPARANASE-LIKE PROTEIN HPA2C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20483645; PubMed=11027606;
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,
RA Hircok M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;
RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian
Heparanase Family Member."
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000). *OKS*
DR EMBL; AF282887; AAG23423.1; -.
SQ SEQUENCE 592 AA; 66580 MW; 95C384AD9A74258E CRC64;

Query Match 98.6%; Score 2746; DB 4; Length 592;
Best Local Similarity 90.2%; Pred. No. 1.3e-212;
Matches 534; Conservative 0; Mismatches 0; Indels 58; Gaps 1;

QY 1 MRVLCAPPEAMPSSNSRPPACIAPGALYALLHLSLSSQAGDRRLPVDRAAGLKEKTL 60
DB 1 MRVLCAPPEAMPSSNSRPPACIAPGALYALLHLSLSSQAGDRRLPVDRAAGLKEKTL 60
QY 61 ILDDYSTKNPVRTVNEFLSLQDPSIHDGWLDFLSKRLVTLARGLSPAFLRFGGKRT 120
DB 61 ILDDYSTKNPVRTVNEFLSLQDPSIHDGWLDFLSKRLVTLARGLSPAFLRFGGKRT 120
QY 121 DFLOFQNLRLNPAKSRGGPGDYILKNYEDDIVRSDVALDKGCKIAHQHPDVMLELQREK 180
DB 121 DFLOFQNLRLNPAKSRGGPGDYILKNYEDDIVRSDVALDKGCKIAHQHPDVMLELQREK 180
QY 181 AAQHLVLLKEQFSNTYSNLILTEPNNTYTMHGRAVNSQLGKDYIQLKSLLOPIRIYSR 240
DB 181 AAQHLVLLKEQFSNTYSNLILTEPNNTYTMHGRAVNSQLGKDYIQLKSLLOPIRIYSR 240
QY 204 -----EPNNYRTMHGRAVNSQLGKDYIQLKSLLOPIRIYSR 242
DB 241 ALSLLKYSASKKYNISWELGNENPNYRTMHGRAVNSQLGKDYIQLKSLLOPIRIYSR 300

QY 243 LYGNIGRPRKNNVIALLDGFMKVAGSTVDVATWQHICYIDGRVVKVMDFLKTRLLDLSQ 302
DB 303 LYGNIGRPRKNNVIALLDGFMKVAGSTVDVATWQHICYIDGRVVKVMDFLKTRLLDLSQ 360
QY 303 IRKQKVVNTYTPGKKIWLEGVVTTSAGGTNNLSDSYAAGFLWNLTLGLMANGIDVIR 362
DB 361 IRKQKVVNTYTPGKKIWLEGVVTTSAGGTNNLSDSYAAGFLWNLTLGLMANGIDVIR 420
QY 363 HSFDHGYNHLVDQNFNPLDYWLSLLYKRLIGPKVLAVHVGAGLQKRPGRVIRDKLRI 422
DB 421 HSFDHGYNHLVDQNFNPLDYWLSLLYKRLIGPKVLAVHVGAGLQKRPGRVIRDKLRI 480
QY 423 YAHCTNHHNNHNVYRGSTITLFIINLHRSRKKIKLAGTLRDKLVHQLVYLLQPYGQGLKSK 482
DB 481 YAHCTNHHNNHNVYRGSTITLFIINLHRSRKKIKLAGTLRDKLVHQLVYLLQPYGQGLKSK 540
QY 483 QLNGQPLVMVDDGTLPELKPRLRAGRTLVIPTVMGFFVKNVNALACRYR 534
DB 541 QLNGQPLVMVDDGTLPELKPRLRAGRTLVIPTVMGFFVKNVNALACRYR 592

RESULT 3
QY 09HB39 PRELIMINARY; PRT; 480 AA.
AC 09HB39;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE HEPARANASE-LIKE PROTEIN HPA2A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20483645; PubMed=11027606;
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,
RA Hircok M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;
RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian
Heparanase Family Member."
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000). *OKS*
DR EMBL; AF282885; AAG23421.1; -.
SQ SEQUENCE 480 AA; 53900 MW; F75F89F67AC1FF83 CRC64;

Query Match 89.0%; Score 2478; DB 4; Length 480;
Best Local Similarity 89.9%; Pred. No. 3.6e-191;
Matches 480; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

QY 1 MRVLCAPPEAMPSSNSRPPACIAPGALYALLHLSLSSQAGDRRLPVDRAAGLKEKTL 60
DB 1 MRVLCAPPEAMPSSNSRPPACIAPGALYALLHLSLSSQAGDRRLPVDRAAGLKEKTL 60
QY 61 ILDDYSTKNPVRTVNEFLSLQDPSIHDGWLDFLSKRLVTLARGLSPAFLRFGGKRT 120
DB 61 ILDDYSTKNPVRTVNEFLSLQDPSIHDGWLDFLSKRLVTLARGLSPAFLRFGGKRT 120
QY 121 DFLOFQNLRLNPAKSRGGPGDYILKNYEDDIVRSDVALDKGCKIAHQHPDVMLELQREK 180
DB 121 DFLOFQNLRLNPAKSRGGPGDYILKNYEDDIVRSDVALDKGCKIAHQHPDVMLELQREK 180
QY 181 AAQHLVLLKEQFSNTYSNLILTEPNNTYTMHGRAVNSQLGKDYIQLKSLLOPIRIYSR 240
DB 150 -----EPNNYRTMHGRAVNSQLGKDYIQLKSLLOPIRIYSR 186
QY 241 ASLYGPNIGRPRKNNVIALLDGFMKVAGSTVDVATWQHICYIDGRVVKVMDFLKTRLLDLS 300
DB 187 ASLYGPNIGRPRKNNVIALLDGFMKVAGSTVDVATWQHICYIDGRVVKVMDFLKTRLLDLS 246
QY 301 DOIRKIQVNTYTPGKKIWLEGVVTTSAGGTNNLSDSYAAGFLWNLTLGLMANGIDV 360
DB 247 DOIRKIQVNTYTPGKKIWLEGVVTTSAGGTNNLSDSYAAGFLWNLTLGLMANGIDV 306

```
QY 361 IRHSFDDHGYNHLVDQNFPLPDYWLSSLYKRLIGPKVLAVHVGAGLQRPGRVIRDKL 420
|||||
Db 307 IRHSFDDHGYNHLVDQNFPLPDYWLSSLYKRLIGPKVLAVHVGAGLQRPGRVIRDKL 366
|||||
QY 421 RIYAHCTNNHHNNYVRSSTLFTLINLHRSRKKTKLAGTLRDKLVHQLYLLQPYGOEGLSK 480
|||||
Db 367 RIYAHCTNNHHNNYVRSSTLFTLINLHRSRKKTKLAGTLRDKLVHQLYLLQPYGOEGLSK 426
|||||
QY 481 SVLQNGQPLVMVDGTLPELKPRLRAGRTLVIPTVMGFFVVKNNALACRYR 534
|||||
Db 427 SVLQNGQPLVMVDGTLPELKPRLRAGRTLVIPTVMGFFVVKNNALACRYR 480
|||||
RESULT 4
Q9QZF8 PRELIMINARY; PRT; 536 AA.
AC Q9QZF8;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE HEPARANASE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Podyma K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;
RT "Heparanase from parathyroid cell line.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF184967; AAF04563.1; -.
SQ SEQUENCE 536 AA; 60569 MW; 6208B1FD9EE28421 CRC64;

Query Match 34.2%; Score 953; DB 11; Length 536;
Best Local Similarity 39.3%; Pred. No. 3e-68;
Matches 216; Conservative 83; Mismatches 165; Indels 86; Gaps 10;

QY 42 GDRRLPVDRAAG-LKEKTLILLDVSTKNPVRTVNENFSLQLODPSIIHD-GWLDLFSK 99
|||
Db 12 GLRLALTOGTPAGTAKDQVDLEFFTKRLFQSVSPSFLSITDASLATDPRFLTGLSP 71
|||
QY 100 RLVTLAGLSAPLREGGKRTDFLQFONLNPNAKSGGPGPDYLYKNYEDIVRSVDALD 159
|||
Db 72 RLRLARGLSPAYLREGGKRTDFLIF-----DPNKEPTSEERSYMQSDNDICGSE--- 123
|||
QY 160 KQGGCKIAQHPVMELOREKAAQMHVLLKQF-----SNTYS----- 198
|||
Db 124 -----RVSADVLRLQMEWPFQ-ELLLREYQREFKNSTYSRSSVDMLYSFAKCSRL 175
|||
QY 199 -----NLIL-----TEPNYRTMHGRAVNGSQL 221
|||
Db 176 DLIFGLNALLRTPDLRWNSSNAQLLNYCSSKGYNTISWELGNPNFQWKAQISIDGLQL 235
|||
QY 222 GKDYIQLKSLLOPIRYSRASLYGPNIGRPKNVIALLDGFMKAVGSTVDVATWQHCYID 281
|||
Db 236 GEDFVELHKLQK-SAFQNAKLYGPDIGQPGKTVKLLRSFLKAGGVEIDSLTWHHYLN 294
|||
QY 282 GRVVKVMDFLKRLTLDTLSQIRKIQKQVNTYTPGKKIWLEGVVTYSAGGTNNLSDSYAA 341
|||
Db 295 GRVATKEDFLSDVDTFTLSVQKILKVTKEMTPGKKVWLGETSSAYGGGAPLLSNTFAA 354
|||
QY 342 GFLWNLATGLMANQGDVIRHSFFDGHYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAV 401
|||
Db 355 GFNWLKGLSAQLGIEVYVNRQVFFGAGNYHLVDENFEPLDYLWLSLLFKLGPVKVMS 414
|||
QY 402 HVAGLQRPGRVIRDKLRIYAHCTNNHHNNYVRSITLFIINLHRSRKKIKLAGTLRD 461
|||
Db 415 RVKGPD-----RSKRLVYLHCTNVYHPRYREGDGLTVLVNLHNVTKHLKLPDPMFS 465
|||
QY 462 KLVHQLYLLQPYGOEGLSKSVQLNGQPLVMVDGTLPELKPRLRAGRTLVIPTVMGFF 521
|||||
```

```
Db 466 RPVDKYLKFGSDGLLSKSVQLNGQTKMVDQTLTPALTEKPLPAGSSLSVAFSYGFF 525
|||||
QY 522 VKNNVNALAC 531
|||
Db 526 VIRNAKIAAC 535
|||||
RESULT 5
Q9UL39 PRELIMINARY; PRT; 545 AA.
AC Q9UL39;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HEPARANASE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=20229546; PubMed=10764835;
RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;
RT "Heparanase expression in invasive trophoblasts and acute vascular
RT damage.";
RL Glycobiology 10:467-475(2000).
DR EMBL; AF084467; AAD54516.1; -.
SQ SEQUENCE 545 AA; 61418 MW; 67B80ACD73C5A9A1 CRC64;

Query Match 33.6%; Score 937; DB 4; Length 545;
Best Local Similarity 38.1%; Pred. No. 5.9e-67;
Matches 219; Conservative 78; Mismatches 184; Indels 94; Gaps 11;

QY 17 RPPACIAPGALYIALLLHLSLSSQAGDRRLPVDVRAAGLKEKTLILLDVSTKNPVRTVNE 76
|||
Db 4 RSKPALPPPLMLLILGLPLGPLSPGAPRAQA-----QDDVVDLDFTQEPHLVSP 56
|||
QY 77 NELSLOLDPSIIHD-GWLDLFSKRLVTLARGLSAPLRFEGKRTDFLQFONLNPNAKSR 135
|||
Db 57 SFLSVTIDANLADTPREFLILGSPKRLARGLSPAYLRFEGGKRTDFLIF-----DPKES 112
|||
QY 136 GGPDPYLYKNYEDIVRSVDALDQKQCKIAQ-HPDVMLELOREKAAQMHVLLKQF- 193
|||
Db 113 TFEERSYMQSQVQNDI-----CKYGSIPDPVEEKLREWPYQEQ-L-LLREHYQ 159
|||
QY 194 -----SNTYS-----NLIL----- 202
|||
Db 160 KFKKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTPDLRWNSSNAQLLIDYCCKGYN 219
|||
QY 203 -----TEPNYRTMHGRAVNGSQLGKDYIQLKSLLOPIRYSRASLYGPNIGRPKNVI 256
|||
Db 220 ISWELGNPNFSLKADIFINGSQLGDFIQLHKLRLK-STFKNAKLYGPDVGPQRKTA 278
|||
QY 257 ALLDGMKAVGSTVDVATWQHCYIDGRVVKVMDFLKRLTLDTLSQIRKIQKQVNTYTPG 316
|||
Db 279 KMLKSLKAGGEVDSVTWHYLYNGRTATREDFLNPDPDLDFISSVQKQVQVVESTRPG 338
|||
QY 317 KKIWLEGVVTYSAGGTNNLSDSYAAAGFLMNTLGMANQGDVIRHSFFDGHYNHLVDQ 376
|||
Db 339 KKVWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGTEVVMQVFFGAGNYHLVDE 398
|||
QY 377 NEPLPDYWLSSLYKRLIGPKVLAVHVGAGLQRPGRVIRDKLRIYAHCTNNHHNNYVR 436
|||
Db 399 NEPLPDYWLSSLYKRLIGPKVLAVHVGAGLQRPGRVIRDKLRIYAHCTNNHHNNYVR 449
|||
QY 437 GSITLFIINLHRSRKKIKLAGTLRDKLVHQLYLLQPYGOEGLSKSVQLNGQPLVMVDGT 496
|||
Db 450 GDULTYAINLHNVTKYLRYPYPSNKKQVQDYLLRPLGPHGLLSKSVQLNGLTLKQVDDQT 509
|||
QY 497 LPELKPRLRAGRTLVIPTVMGFFVVKNNALAC 531
|||||
```

[illegible]

```
Db 218 YNISWELNGENSPNSFORAKGIFINGRQLGEDFIEFRKLGLK-SAFKNAKLYGPDIGQPRRN 276
Qy 255 VIALIDGPMKVAGSTVDAVTWQHCVYIDGRVVKVMDFLKRLDLSDOIIRKIQKVNTYT 314
Db 277 TVKMLKSLFKAGGEVDSVTWHYVNGRIATKEDFLNPDLIDTFISSVQKTLRIVEKIR 336
Qy 315 PGKIKWLEGVVTTSAGGPNNSDSYAAGFLWNTLGLMANGIDVIRHSFFDHGYNHLV 374
Db 337 PLKKVWLGETSFAFGGAPFLSNTFAAGFMWLDKLGLSARMGIEVVMRQVLFAGNYHLV 396
Qy 375 DONENPLDYWLSLYKRLGPKVLAVHAGLQRPGRVIRDKLRIYAHCTNHNHNY 434
Db 397 DGNFEPLDYWLSLFLKLVGNKVLMSVKGPD-----RSKFRYLHCTNKHPRY 447
Qy 435 VRGSITFLIINLHRSKKIKLAGTLRDLKLVHQLQPYGOEGLKSKSVQLNGQPLVMVDD 494
Db 448 KEGDLTLVALNLHNVTKHLEPHLFENKQVDKYLKPSGTDGLSKSVQLNGQILKWDE 507
Qy 495 GTLPDLKPRPLRAGRTLIVPPVTMGFFVVKVNNALAC 531
Db 508 QTLPALTEKPLHPGSLGMPFSGFFVIRNAKVAAC 544

RESULT 8
Q90YK5 ID Q90YK5 PRELIMINARY; PRT; 523 AA.
AC Q90YK5
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE HEPARANASE.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=11387326;
RA Goldshmidt O., Zcharia E., Alingorn H., Guatta-Rangini Z., Atzmon R.,
RA Michal I., Pecker I., Mitrani E., Vlodavsky I.;
RT "Expression Pattern and Secretion of Human and Chicken Heparanase Are
RT Determined by Their Signal Peptide Sequence.";
RA J. Biol. Chem. 276:29178-29187(2001).
DR EMBL; AY037007; AAK82648.1; -.
SQ SEQUENCE 523 AA; 58386 MW; 8EB0B7B18C9BF881 CRC64;

Query Match 30.0%; Score 836.5; DB 13; Length 523;
Best Local Similarity 35.7%; Pred. No. 6.9e-59;
Matches 200; Conservative 79; Mismatches 187; Indels 95; Gaps 10;

Qy 27 LYLALLHLSSQAGDRRPLPVDRAAGLKEKTLILLDVTKNPVRTVNEFLSLQLOPS 86
Db 1 MLVLLLVLLIA-----VPPRTAEIQ-----LGLREPIGAVSPAFSLTLDAS 44
Qy 87 ITHD-GWLDLSSKRLVTLARLFGGKRTDFLQFQNLNRPKSRGGPGPDYILK 145
Db 45 LARDPRFVALLRHPLKHTLASGLSPGLRFGTSTDFLIF-----NPNKD-----S 90
Qy 146 NYEDIVRSVDALD-----KQKCKIAQHP-----D 171
Db 91 TWEERVLSEFOAKDVCEAWPSFAVPKLLLTQWPLQELKLLLAHSHWKHKNTITRSTLD 150
Qy 172 VM-----LELQREKAAQMLHVLVLLKEQFSNTYSNLIITEPNNTYT 210
Db 151 ILHTPASSGGFLVFLGNALLRRAGLQWDSSNAKOLLGYCAQORSYNISWELGNEPNSFRK 210
Qy 211 MHGRAVNSQLCKDYIQLSLQIPRIYSRASLYGPNIGRPNKNTVALLDGMFKVAGSTV 270
Db 211 KSGICIDFQLGRDPRVHLRQLLSQHPLEHAELYGLDVGQPKHTQHLLRSFMKSGGKAI 270
Qy 271 DAVTWQHCVYIDGRVVKVMDFLKRLDLSDOIIRKIQKVNTYTTPGKKIWLGVVTTSG 330
```

```
Db 271 DSVTWHYVNGRSATREDFLSPEVLDSFATAIHDLVGLVEATVPCKKWLGETGSAYGG 330
Qy 331 GTNNLSDSYAAGFLWNTLGLMANGIDVIRHSFFDHGYNHLVDOONFPLPDYWLSSLY 390
Db 331 GAQLSNTYVAGFMWLDKLGLAARRGIDVVMRQVSGAGSYHLVDAGFKPLPDYWLSSLY 390
Qy 391 KRLIGPKVLAVHAGLQRPGRVIRDKLRIYAHCTNHNHNYVRGSIITFLIINLHRSR 450
Db 391 KRLVGRVLQASVEQADAR-RP-----RVYLCTNPRHPKYREGDVTLEFALNSNT 441
Qy 451 KKKIAGTLRDLKLVHQLQPYGOEGLKSKSVQLNGQPLVMVDDGTLPDLKPRPLRAGRT 510
Db 442 QSLQPLPKQLWKSVDQYLLPHGKDSILSREVQLNGRLQLMVDDDETLPALHEMALPGST 501
Qy 511 LVIPPVTMGFFVVKVNNALAC 531
Db 502 LGLPAFSYGFYIRNAKAIAC 522

RESULT 9
Q9SDA1 ID Q9SDA1 PRELIMINARY; PRT; 521 AA.
AC Q9SDA1
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE HYPOTHETICAL 57.8 KDA PROTEIN.
GN F13G24.30.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Van der Schueren J., Chuang Y.J., Voet M., Robben J.,
RA Volckaert G., Bancroft I., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL133421; CAB62595.1; -.
DR InterPro; IPR001254; Trypsin.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 521 AA; 57831 MW; 07D8664A4B305CC2 CRC64;

Query Match 10.8%; Score 302; DB 10; Length 521;
Best Local Similarity 24.8%; Pred. No. 7.7e-16;
Matches 139; Conservative 73; Mismatches 183; Indels 166; Gaps 27;

Qy 75 NENFLSLQID--PSIIHD-----GWLDLFS-----SKRLVTLA-RGLSPAFLEGGKR 119
Db 22 DENFVCATLDWNP---HDKCNVDQCQWGYSSVINMDLTRPLLTAKAKFKPLRIRIGSSL 78
Qy 120 TDFLOFQ--NLNRP-----KSRGGCPDYIYLNKYEDDIIVRSDVALDKQKCKIAQHPDM 173
Db 79 QDOVIYDVGNLKTPCRPFOKMNSG-----LFGFSKGLHMKRWD-- 117
Qy 174 LELQREKAAQMLHV-----LLKEQFSNTYSNLIITEPNNTYRTMHGRAVN----- 217
Db 118 -EENSELTATGAVVTFGLNALGRHKLKGANGGADHINTQDFLNTYVSKGVIVDSWEF 176
Qy 218 GSQL-----GKDYIQLKSLQLPIRIYSRASLYGPNIGRPNKNTVALLDGF-- 262
Db 177 GNELSSGGVSGVSAELYKGLDILVKDVIN--KVYKNSWLHKKPILVAP-----GGFY 226
Qy 263 -----MKVAG-STVDVATWQHCVIDGR-----VVKVMDFLKRLDLSDOIIRKIQ 307
Db 227 EQQWTKLEISGSPVVDVVT-HHIYINLGSNDPALVKKIMD---PSYLSQVSKTEFDVN 282
```

```
QY 308 KVVNTYTPGKKIWLGVVTTTSG-----GTNNLSDSYAAGFLWNLTLGLMLANQGDVVIRH 363
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 283 QTIOEHGP-----WASPMWGESGAYNSGGRHVSDTFIDSFYLDQLGMSARHNTKVYCRQ 338
QY 364 SFEDHGYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHVAGLQKRPGRVIRDKLRIY 423
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 339 TLVGFYGLLEKGTVPNPDIYSALLWRLMGKGLAVQTDG-----PPQLRVY 387
QY 424 AHCTNHHNHNHYVRSITLFIINLH-----RSRKKIKLAGTLRDKLV 464
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 388 AHCSKG-----RAGVTLLINLSNQSDFTVSVNSGINVWNAESRKKSLDLTLKRPFS 441
QY 465 -----HOYLLOPYQOEG-LKSKSVOLNQPLVMVDDGTLPELKPRLPRAGRT 510
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 442 WIGSKASDGYLNREYHLTP--ENGVLRSKTMVLNGKSLKPTATGDIPSLEPVRSVNSP 499
QY 511 LVIPPTMGFFVWKNVNALAC 531
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 500 LNVPLMSFIVLPNFDASAC 520
```

```
RESULT 10
Q9FF10 PRELIMINARY; PRT; 543 AA.
AC Q9FF10;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SIMILARITY TO HEPARANASE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=97471969; PubMed=9330910;
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
RT features of the 1.6 Mb regions covered by twenty physically assigned
RT p1 clones.";
RL DNA Res. 4:215-230(1997).
DR EMBL; AB005249; BAB03947.1; -.
DR InterPro; IPR001254; Trypsin.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
SQ SEQUENCE 543 AA; 60250 MW; 0FA2248948282FF6 CRC64;
```

```
Query Match 10.8%; Score 302; DB 10; Length 543;
Best Local Similarity 24.8%; Pred. No. 8.2e-16;
Matches 139; Conservative 73; Mismatches 183; Indels 166; Gaps 27;

QY 75 NENFLSLQLD--PSIITHD-----GWLDFLS---SKRLVTLA-RGLSPAFLRPGGRK 119
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 44 DENFVCATLDWNP---HDKCNVDQCPWGYSSVINMDLTPLLTKAKAFKPLRIGGSL 100
QY 120 TDFLQFO--NLRNPA----KSRGGPGPDYLYKNYEDDIVRSDVALDKQCKIAQHPDYM 173
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 101 QDOVIYDVGNLKTPCRPFQKMNSG-----LFGFSKGCLHMKRWD-- 139
QY 174 LELQREKAQHMLV-----LLKQESNTYSNLLITEPNNYRTMHGRAVN---- 217
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 -ELNSFLTATGAVTFCGLNALRGRHRLCKAMGGAWDHINTQDFLNTYVSKGVIVDSWEF 198
QY 218 GSOL-----GKDYIQLKSLLOPIRIYGRASLYGNIGRPRKNVIALDGF- 262
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 199 GNELSGVGASVSAELYGKDLVLKDVIN--KVYKNSWLHKPILVAP-----GGFY 248
QY 263 -----MKVAG-STVDVATWQHCHYIDGR-----VVKVMDFLKTRLLDLSQTRKIO 307
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Db 249 EQQWYTKLLISGPSVVDVVT-HHIYNLGSNDPALVKKIMD---PSYLSQVSKTFKDVN 304
QY 308 KVVNTYTPGKKIWLGVVTTTSG-----GTNNLSDSYAAGFLWNLTLGLMLANQGDVVIRH 363
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 305 QTIOEHGP-----WASPMWGESGAYNSGGRHVSDTFIDSFYLDQLGMSARHNTKVYCRQ 360
QY 364 SFEDHGYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHVAGLQKRPGRVIRDKLRIY 423
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 361 TLVGFYGLLEKGTVPNPDIYSALLWRLMGKGLAVQTDG-----PPQLRVY 409
QY 424 AHCTNHHNHNHYVRSITLFIINLH-----RSRKKIKLAGTLRDKLV 464
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 410 AHCSKG-----RAGVTLLINLSNQSDFTVSVNSGINVWNAESRKKSLDLTLKRPFS 463
QY 465 -----HOYLLOPYQOEG-LKSKSVOLNQPLVMVDDGTLPELKPRLPRAGRT 510
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 464 WIGSKASDGYLNREYHLTP--ENGVLRSKTMVLNGKSLKPTATGDIPSLEPVRSVNSP 521
QY 511 LVIPPTMGFFVWKNVNALAC 531
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 522 LNVPLMSFIVLPNFDASAC 542
```

```
RESULT 11
Q9FLK8 PRELIMINARY; PRT; 516 AA.
AC Q9FLK8;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SIMILARITY TO HEPARANASE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=98290546; PubMed=9628582;
RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
RT Sequence features of the regions of 1,456,315 bp covered by nineteen
RT physically assigned p1 and TAC clones.";
RL DNA Res. 5:41-54(1998).
DR EMBL; AB010073; BAB08480.1; -.
DR InterPro; IPR001917; Aminotransf.2.
DR PROSITE; PS00599; AA_TRANSFER_CLASS_2; UNKNOWN_1.
SQ SEQUENCE 516 AA; 57414 MW; 6943E6D3D89FB93C CRC64;

Query Match 9.4%; Score 261.5; DB 10; Length 516;
Best Local Similarity 21.1%; Pred. No. 1.4e-12;
Matches 121; Conservative 94; Mismatches 201; Indels 157; Gaps 22;

QY 57 EKTLLILDVSTKNPVRTVNEFLSLQLD-----PSIIHDGWLDFLSKR 100
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2 ERTTLVDGSR--TAETDENFICATLDWWPEKCNQCWPWGYASLINL--LASPL 55
QY 101 LVTLAGLSAPFLRFGGKRTDFLQFO--NLRNPAKSRGGPGPDYLYKNYEDDIVRSDVAL 158
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 56 LAKAIQAFRTLIRIGSGSLQDQVIYDVGLKTPCTQ-----FKTDDGL-----F 100
QY 159 DKQKCKIAQHPDVMLELQREKAA---QMLVLVLLKEQFSNT-----YSNLITLTPNNYR 209
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 101 GFSEGLLYMKRWDENVNHFENATCAIVTFCGLNALHGNKLNGLTAWGDMWDTHTQDFMNT 160
QY 210 TMHGRAVN---GSQL-----GKDYIQLKSLLOPIRIYGRASLYGNIGR 251
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 161 VSKGYAIDSWERGNELSGSIWASVVELYKDLVLKVNKNVKNRSRT----- 210
QY 252 RKNVIALLDGF-----MKVAGSTVDVATWQHCHYIDG-----RVVKVMDFLKTRLL 296
```


Db 211 -KPLVAPGFFBQWSELSURLSGPGVLDVLTTHIYNLPGNDPKLVNKILD---PNYL 266
QY 297 DTLSDQIRKIOKVNTYTPGKKIWLGVVTTSGAGTNN-----LSDSYAAGFLWLNTLGM 351
Db 267 SGISELFANVQTIQEHCPWAAWV---GEAGGAFNSGGQVSETFINSFWYLDQLGI 321
QY 352 LANGIDVVIHRSFFDHCYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHAGLQKRP 411
Db 322 SSKHNTKYICQALVGGFYGLLETFEPNPDYYSALLWHLRMLGKILGVQTTA----- 375
QY 412 PGRVIRDKLRIYAHCTNHNHNHYVSGSTLFLINLHR-----SRKK 452
Db 376 -----SEVLRAVHCXKR-----RAGTILLINLSKHTTFTVAVSNGVKVYLQAESMKR 424
QY 453 IKLAGTLDKLV-----HOYLQPYGEG-LKSKSVOLNGOPLVMVDDGTLP 498
Db 425 KSFLETIKSVKSVGNKASDGYLNREEVHLSF--KQGLRSKIMLLNGKPLVPATGDIP 482
QY 499 ELKPRPLRAGRTLVIPTVMGFFVKKVNALAC 531
Db 483 KLEPVRHGKSPVYINPLSISFVLPTFDAPAC 515

RESULT 12

Q9LR08 PRELIMINARY; PRT; 527 AA.
AC Q9LR08;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BETA-GLUCURONIDASE.
GN SGUS.
OS Scutellaria baicalensis.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Lamiales; Lamiaceae; Scutellaria.
OX NCBI_TaxID=65409;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=20418130; PubMed=10858442;
RX Sasaki K., Taura F., Shoyama Y., Morimoto S.;
RA "Molecular Characterization of a Novel beta-Glucuronidase from
RT Scutellaria baicalensis Georgi.;
RL J. Biol. Chem. 275:27466-27472(2000).
DR EMBL; AB040072; BAA97804.1; -
DR InterPro; IPR001179; FKBP_PPIase.
DR PROSITE; PS00453; FKBP_PPIASE_1; UNKNOWN_1.
SQ SEQUENCE 527 AA; 58772 MW; A5DE7C423F2A1E2B CRC64;

Query Match 9.2%; Score 256.5; DB 10; Length 527;
Best Local Similarity 24.2%; Pred. No. 3.6e-12;
Matches 129; Conservative 81; Mismatches 207; Indels 115; Gaps 27;
QY 56 KEKTLILDVSTKNPRTVNENFSLQLD---PSIIHDG---W-----LDFLSSKRLV 102
Db 26 EETTVIKIE---ENPVAQTIDENYVCATLDLWPPTKCNYGCPWCKSSPLNLD-LNNNIIR 81
QY 103 TLARGLSPAFLRFGGKRTDFLQFNLRNPAKSRGGPGDYVLKNYEDDIVSDVALDKQK 162
Db 82 NAVEFAPLKLRFGGTQDLRLVIQT-----SREPCDSTFYN-----TNLIDFSH 128
QY 163 GKCTIAQHPDV-----LELQREKAAQMHVLVLLKEQ-----FSN 195
Db 129 ACLSLDRWDEINFILETGEAVFGLNALRGKTVETKGIKIDGGYLGTTTAVGEWDYSN 188
QY 196 TYSNLIITEPNNTMTMGRVNGSQLG-----KDYI-QKSLQPIR-IYSRASL 243
Db 189 SKFLIEYSLKGYKHIRGWTL-GNELGHTLFIGVSPEDYANDAKKLHELVEIKYQDQGT 247
QY 244 YGPNIGRPRKNVIALDGFEMKVAQSTVDA-VTWQHCYIDGRVVKVMDFLKRLD-TLSD 301

Db 248 M-PLIAP--GAIFDLEWYTEFIDRTPELHVATHHMYNLGSGD--DALKOVLLTASFDD 302
QY 302 QIRK-----IOKVNTYTPGKK--IWLEGVVTTSAGGTNNLSDSYAAGFLWLNTLGMLAN 354
Db 303 EATKSMYELQKIVN--RPGTKAVAWIGEAGGAFNSGGQISNTFINGFWYLNMLGYLSAL 360
QY 355 QGIDVVIHRSFFDHCYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHAGLQKRP 414
Db 361 LDTKTFCQTLTGNGYGLLQGTGTYIPNPDYYSALLWHLRMLGSKVLKTEIVGTFK----- 413
QY 415 VIRDKLRIYAHCTNHNHNHYVSGSTLFLINLHRSKKIKLA-----GTLRDKLVHQLV 468
Db 414 ----NVYIYAHCAKSN-----GITMLVLN-HDGESSVKISLDESKYGSKRE----EYH 458
QY 469 LQPYQEGGLKSKSVOLNGOPLVMVDDGTLPKLPRLRAGRTLVIPTVMGTF 520
Db 459 LTPV-NNNLQSELVXLNGELLHLDPSGVI PALNPVEKDNSKOLEVAPISFMF 509

RESULT 13

Q9FZP1 PRELIMINARY; PRT; 536 AA.
AC Q9FZP1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SIMILARITY TO HEPARANASE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-COLUMBIA;
RC Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RL "Structural analysis of Arabidopsis thaliana chromosomes 5. XI. ";
RT Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB028613; BAB10787.1; -
SQ SEQUENCE 536 AA; 59654 MW; 24202B8E82F3DB0E CRC64;

Query Match 8.8%; Score 246; DB 10; Length 536;
Best Local Similarity 21.4%; Pred. No. 2.6e-11;
Matches 119; Conservative 86; Mismatches 195; Indels 156; Gaps 25;

QY 71 VRTVNEFSLQLD---PSIIHDG---WLDFLSSKRL-----VTLARGLSPAFLRFGG 117
Db 41 VGTIDEDEFICATLDWWPPEKCDYGSCSW-DHASILNLDLTIIVVPIKNTAFAPLIRIGG 99
QY 118 KRTDFLQFNLRNPAKSRGGPGDY-----YLKNYEDDIVSDVALDKQCKIAQHPD 171
Db 100 TLQDIIIVET-----PDSKQCLPFTKN-----SSILFGYTGQCLPMRRRD 140
QY 172 VMLQREKAAQMHVL-----LKE-----QFSNTYSNLIITEPNNY----- 209
Db 141 ELNAFERKTGTVIFGLNALSGRSIKSNGEATGANNYNAESFIRTAENNYTIDGWELG 200
QY 210 -----TMHGRAVNGSQLGDYITQLKSLQPIRIYGRAS-----LYGPNIGRPRKNVIALD 260
Db 201 NELCGSGVARGVANGAYAITDITNLRNIVN--RVYKNVSPMLVIVGPG----- 245
QY 261 GFMKV-----AGSTVDVAVTWOHCY-----IDGRVVKVMDFLKTRLIDFLSDQIRKI 306
Db 246 GFEFVDWTEYLKAEINLKNATT-RHIYDLGPGVDEHLIE--KILNPSYLDQEAQSFSL 302
QY 307 QKVNTYTPGKKIWLGVVTTSG-----GTNNLSDSYAAGFLWLNTLGMLANQGDIVVIR 362
Db 303 KNIKNSSTKAVAW-----VGESGGAYNSGRNLVSNAFVYFWYLDQLGMASLYDTKTYCR 358
QY 363 HSFFDHGYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHAGLQKRPGRVIRDKLRI 422

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:10:18 ; Search time 53.59 Seconds
(without alignments)
243.390 Million cell updates/sec

Title: US-09-836-461-2
Perfect score: 2785
Sequence: 1 MRVLCAPFAMPSSNRPPA.....PVTMGFFVKNVNALACRYR 534

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	936.5	33.6	543	2	US-08-922-170B-10
2	936.5	33.6	543	4	US-09-071-739B-2
3	936.5	33.6	543	4	US-09-260-038B-2
4	932.5	33.5	543	4	US-09-181-336-13
5	889.5	31.9	532	4	US-09-181-336-15
6	805	28.9	380	4	US-09-181-336-19
7	779	28.0	380	4	US-09-181-336-17
8	94.5	3.4	685	2	US-08-878-989-1
9	94.5	3.4	685	3	US-09-136-282-2
10	94.5	3.4	685	4	US-09-272-796-1
11	94.5	3.4	685	4	US-09-505-744-2
12	93.5	3.4	934	1	US-08-215-805A-80
13	93	3.3	471	1	US-08-257-341-9
14	91.5	3.3	915	4	US-09-346-237-2
15	91.5	3.3	928	4	US-09-514-599-4
16	91	3.3	1788	2	US-08-962-284-2
17	89.5	3.2	1118	1	US-07-934-393B-2
18	89.5	3.2	1118	1	US-08-278-089A-2
19	89.5	3.2	1118	2	US-08-838-957A-2
20	89	3.2	632	4	US-09-016-080-1
21	89	3.2	1280	2	US-08-583-276-19
22	89	3.2	1280	6	5206352-4
23	88	3.2	1106	1	US-08-435-675B-5
24	88	3.2	1106	1	US-08-336-257A-8
25	87.5	3.1	752	1	US-08-281-193-2
26	87.5	3.1	752	1	US-08-422-106-2
27	87.5	3.1	752	2	US-08-735-716-2

28	87.5	3.1	752	2	US-08-555-568B-2	Sequence 2, Appli
29	87.5	3.1	752	4	US-09-519-223-2	Sequence 2, Appli
30	87.5	3.1	752	5	PCT-US95-08069-2	Sequence 2, Appli
31	87.5	3.1	834	3	US-08-539-205A-6	Sequence 6, Appli
32	87.5	3.1	1122	1	US-08-278-089A-6	Sequence 6, Appli
33	87.5	3.1	1122	2	US-08-838-957A-6	Sequence 6, Appli
34	87	3.1	1124	1	US-08-323-474-2	Sequence 2, Appli
35	87	3.1	1124	5	PCT-US93-06093-2	Sequence 2, Appli
36	86	3.1	1103	1	US-08-455-543A-53	Sequence 53, Appli
37	86	3.1	1103	2	US-08-223-305C-53	Sequence 53, Appli
38	85	3.1	673	4	US-09-078-347A-2	Sequence 2, Appli
39	85	3.1	1086	6	5386025-8	Patent No. 5386025
40	85	3.1	1280	2	US-08-752-447-2	Sequence 2, Appli
41	84.5	3.0	434	2	US-08-989-925-1	Sequence 1, Appli
42	84.5	3.0	484	3	US-09-080-044-7	Sequence 7, Appli
43	84.5	3.0	533	1	US-08-220-151-15	Sequence 15, Appli
44	84.5	3.0	533	1	US-08-413-118-15	Sequence 15, Appli
45	84.5	3.0	533	3	US-08-473-446-15	Sequence 15, Appli

ALIGNMENTS

RESULT 1
US-08-922-170B-10
; Sequence 10, Application US/08922170B
; Patent No. 5968822
; GENERAL INFORMATION:
; APPLICANT: Iris Pecker, Israel Vlodavsky and Elena
; APPLICANT: Feinstein
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A POLYPEPTIDE
; TITLE OF INVENTION: HAVING HEPARANASE ACTIVITY AND EXPRESSION OF
; TITLE OF INVENTION: SAME IN TRANSFUSED CELLS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Robert Sheinbein
; STREET: 2940 Birchtree lane
; CITY: Silver Spring
; STATE: Maryland
; COUNTRY: United States of America
; ZIP: 20906
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; OPERATING SYSTEM: Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted to
; SOFTWARE: an ASCII file
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/922.170B
; FILING DATE: 2 SEP 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedmam, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 910/1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 972-3-5625553
; TELEFAX: 972-3-5625554
; TELEX:
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-922-170B-10

Query Match 33.6% ; Score 936.5 ; DB 2 ; Length 543;

```

; Filing Date: September 2, 1997
; Application Number:
; Filing Date:
; Attorney/Agent Information:
; Name: Friedmann, Mark M.
; Registration Number: 33,883
; Reference/Docket Number: 910/5
; Telecommunication Information:
; Telephone: 972-3-5625553
; Telefax: 972-3-5625554
; Telex:
; Information for Seq ID No. 2:
; Sequence Characteristics:
; Length: 543
; Type: amino acid
; Strandedness: single
; Topology: linear
; US-09-071-739B-2

Query Match 33.6%, Score 936.5; DB 4; Length 543;
Best Local Similarity 37.8%; Pred. No. 3e-91;
Matches 217; Conservative 78; Mismatches 168; Indels 111; Gaps 12;

Qy 18 PPACIAPGALYLALLHLSSLSSQAGDRPLPVDRAGLKEKTLILLDYSTKPNVTVNEN 77
Db 20 PLGLSPGAL-----PRPA-----QAQDVLDLDFFTQEPHLVSPS 55

Qy 78 FLSLQDPSIHD-GWLDLFLSKRLVTLARGLSPAFLREGGKRTDFLOFQNLNPAKSRG 136
Db 56 FLSVTIDANLATDPRFLILLGSPKRLTARGLSPAYLRFPGTKTDFLIF----DPKKEST 111

Qy 137 GPGPDYILKNYEDDVRSDVALDKQKCKIAQ-HPDVMLELQREKAAQHMLVLLKEQF-- 193
Db 112 FEERSYWSQVQNDI-----CKYGSIPDPVEEKLRLWPYQEQ-L-LLREHYQK 158

Qy 194 ---SNTYS-----NLIL----- 202
Db 159 KFKNSTYSSRSDVLYTFANGSGLDIFGLNALLRTADLQWNSSNAQLLLDYCSSKGNYI 218

Qy 203 -----TEPNNYRTMHGRAVNSQLGDKDYIQLKSLLOPIRIYSRASLYGNIGRPRKNVIA 257
Db 219 SWELGNEPNSFLKKADIFINGSQLGEDIYQLHKLLRK-STFKNAKLYGPDVGQPRKTKAK 277

Qy 258 LLDGPMKVAGSTVDVATWQHCVIDGRVVKVMDFLKTRLLDLSQDQIRKIQKVNTYTPGK 317
Db 278 MLKSFLLKAGGEVIDSVTHHHYLLNGRTATREDFLNPDVLDIFISSVQVQVVESTREPK 337

Qy 318 KIWLEGVVTTSAGTNNLSDSYAAGFLMNTLGLMANOGIDVVIRHSFDFHGYNHLVDON 377
Db 338 KVMLETSAYGGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDEN 397

Qy 378 FNPDPYWLSSLYKRLIGPKVLAVHAGLQRPGRVIRDKLRIYAHCTNHHNHYVRG 437
Db 398 FDPDPYWLSSLYKRLIGPKVLAVHAGLQRPGRVIRDKLRIYAHCTNHHNHYVRG 448

Qy 438 SITLFIINLHRSRKIKLAGTLRDKLVHQLYQVQGLSKSVQLNGQPLVWVDDGTL 497
Db 449 DLTLYAINLHNVTYLRPLYPFSPNKQVDKYLRLPLGPHGLLSKSVQLNGLTLKMYDDQTL 508

Qy 498 PELKPRPLRAGRTLVPVPMGFFVVKVNNALAC 531
Db 509 PPLMEKPLRPGSSGLPAPFSYFFVIRNAKVAAC 542

RESULT 2
US-09-071-739B-2
; Sequence 2, Application US/09071739B
; Patent No. 6177545
; GENERAL INFORMATION:
; APPLICANT: Iris Pecker et al.
; TITLE OF INVENTION: HEPARANASE SPECIFIC MOLECULAR PROBES
; TITLE OF INVENTION: AND THEIR USE IN RESEARCH AND MEDICAL
; TITLE OF INVENTION: APPLICATIONS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
; STREET: 20001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Silinote-850TX
; OPERATING SYSTEM: MS DOS version 6.2
; SOFTWARE: Word for Windows version 3.11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,739B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/922,180
```

```

; Filing Date: September 2, 1997
; Application Number:
; Filing Date:
; Attorney/Agent Information:
; Name: Friedmann, Mark M.
; Registration Number: 33,883
; Reference/Docket Number: 910/5
; Telecommunication Information:
; Telephone: 972-3-5625553
; Telefax: 972-3-5625554
; Telex:
; Information for Seq ID No. 2:
; Sequence Characteristics:
; Length: 543
; Type: amino acid
; Strandedness: single
; Topology: linear
; US-09-071-739B-2

Query Match 33.6%, Score 936.5; DB 4; Length 543;
Best Local Similarity 37.8%; Pred. No. 3e-91;
Matches 217; Conservative 78; Mismatches 168; Indels 111; Gaps 12;

Qy 18 PPACIAPGALYLALLHLSSLSSQAGDRRPLPVDRAGLKEKTLILLDYSTKPNVTVNEN 77
Db 20 PLGLSPGAL-----PRPA-----QAQDVLDLDFFTQEPHLVSPS 55

Qy 78 FLSLQDPSIHD-GWLDLFLSKRLVTLARGLSPAFLREGGKRTDFLOFQNLNPAKSRG 136
Db 56 FLSVTIDANLATDPRFLILLGSPKRLTARGLSPAYLRFPGTKTDFLIF----DPKKEST 111

Qy 137 GPGPDYILKNYEDDVRSDVALDKQKCKIAQ-HPDVMLELQREKAAQHMLVLLKEQF-- 193
Db 112 FEERSYWSQVQNDI-----CKYGSIPDPVEEKLRLWPYQEQ-L-LLREHYQK 158

Qy 194 ---SNTYS-----NLIL----- 202
Db 159 KFKNSTYSSRSDVLYTFANGSGLDIFGLNALLRTADLQWNSSNAQLLLDYCSSKGNYI 218

Qy 203 -----TEPNNYRTMHGRAVNSQLGDKDYIQLKSLLOPIRIYSRASLYGNIGRPRKNVIA 257
Db 219 SWELGNEPNSFLKKADIFINGSQLGEDIYQLHKLLRK-STFKNAKLYGPDVGQPRKTKAK 277

Qy 258 LLDGPMKVAGSTVDVATWQHCVIDGRVVKVMDFLKTRLLDLSQDQIRKIQKVNTYTPGK 317
Db 278 MLKSFLLKAGGEVIDSVTHHHYLLNGRTATREDFLNPDVLDIFISSVQVQVVESTREPK 337

Qy 318 KIWLEGVVTTSAGTNNLSDSYAAGFLMNTLGLMANOGIDVVIRHSFDFHGYNHLVDON 377
Db 338 KVMLETSAYGGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDEN 397

Qy 378 FNPDPYWLSSLYKRLIGPKVLAVHAGLQRPGRVIRDKLRIYAHCTNHHNHYVRG 437
Db 398 FDPDPYWLSSLYKRLIGPKVLAVHAGLQRPGRVIRDKLRIYAHCTNHHNHYVRG 448

Qy 438 SITLFIINLHRSRKIKLAGTLRDKLVHQLYQVQGLSKSVQLNGQPLVWVDDGTL 497
Db 449 DLTLYAINLHNVTYLRPLYPFSPNKQVDKYLRLPLGPHGLLSKSVQLNGLTLKMYDDQTL 508

Qy 498 PELKPRPLRAGRTLVPVPMGFFVVKVNNALAC 531
Db 509 PPLMEKPLRPGSSGLPAPFSYFFVIRNAKVAAC 542

RESULT 3
US-09-260-038B-2
; Sequence 2, Application US/09260038B
; Patent No. 6348344
; GENERAL INFORMATION:
; APPLICANT: Maty Ayal-Hershkovitz et al.
; TITLE OF INVENTION: GENETICALLY MODIFIED CELLS AND METHODS FOR
; EXPRESSING RECOMBINANT HEPARANASE
```

AND METHODS OF PURIFYING SAME

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: Mark M. Friedman c/o Anthony Castorina

STREET: 2001 Jefferson Davis Highway, Suite 207

CITY: Arlington

STATE: Virginia

COUNTRY: United States of America

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk

COMPUTER: Twinhead* Slimnote-890TX

OPERATING SYSTEM: MS DOS version 6.2.

SOFTWARE: Word for Windows version 3.11

an ASCII file

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/260, 038B

FILING DATE: 02-Mar-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/071,618

FILING DATE: May 1, 1998

APPLICATION NUMBER: 09/071,739

FILING DATE: May 1, 1998

APPLICATION NUMBER: 08/922,180

FILING DATE: September 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Friedman, Mark M.

REGISTRATION NUMBER: 33,883

REFERENCE/DOCKET NUMBER: 910/16

TELECOMMUNICATION INFORMATION:

TELEPHONE: 972-3-5625553

TELEFAX: 972-3-5625554

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 543

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-260-038B-2

```

RESULT      4
US-09-181-336-13
; Sequence 13, Application US/09181336A
; Patent No. 6242238
; GENERAL INFORMATION:
; APPLICANT: FREEMAN, Craig Geoffrey
; APPLICANT: PARISH, Richard
; APPLICANT: HAMDORF, Brenton James
; APPLICANT: HULETT, Mark Darren
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN
; TITLE OF INVENTION: ENOGLUCURONIDASE AND USES THEREFOR
; FILE REFERENCE: 032505-004
; CURRENT APPLICATION NUMBER: US/09/181.336A
; CURRENT FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: AU PP0062
; EARLIER FILING DATE: 1997-10-28
; EARLIER APPLICATION NUMBER: AU PP0812
; EARLIER FILING DATE: 1997-12-09
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-181-336-13

```

Db 338 KWLGETSSAYGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDEN 397
QY 378 ENPLPDYWLSLLYKRLIGPKVLAVHVGLOKRPGRVIRDKLRIYAHCTNHHNNHYVRG 437
Db 398 FDPPLDYWLSLLFKLVGTKVLMAVSGSKRR-----KLRVYLHCTNTDNPRIYK 448
QY 438 STTLFTINLHRSRKKIKLAGTLRDLKHQYLLQPYGOEGLKSKSVQNLGQPLVMVDDGTL 497
Db 449 DLTVAINLHNVTYKLRPLPYPSNKQVDKYLRLPLGPHGLLSKSVQNLGLTLKMWDDQTL 508
QY 498 PELKPRPLRAGRTL 511
Db 509 PPLMEKPLRPGSSL 522

RESULT 5
US-09-181-336-15
; Sequence 15, Application US/09181336A
; Patent No. 6242238
; GENERAL INFORMATION:
; APPLICANT: FREEMAN, Craig Geoffrey
; APPLICANT: PARISH, Richard
; APPLICANT: HAMDORF, Brenton James
; APPLICANT: HULETT, Mark Darren
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN
; FILE REFERENCE: 032505-004
; CURRENT APPLICATION NUMBER: US/09/181,336A
; CURRENT FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: AU PP0062
; EARLIER FILING DATE: 1997-10-28
; EARLIER APPLICATION NUMBER: AU PP0812
; EARLIER FILING DATE: 1997-12-09
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 532
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-181-336-15

Query Match 31.9%; Score 889.5; DB 4; Length 532;
Best Local Similarity 37.7%; Pred. No. 3.1e-86;
Matches 209; Conservative 75; Mismatches 159; Indels 111; Gaps 12;

QY 18 PPACIAPGALYALALLHLSSLGQAGDRPLPVDRAAGLKEKTLILLDVSTKPNVTVNEN 77
Db 20 PLGLPSGAL-----PRPA-----QAQDVVDLDFFTQBPGLHVS 55
QY 78 FLSLQIDPSTIHD-GWLDLFLSKRLVTLARGLSAPFLREGGKRTDPLQFNLRNPAKSRG 136
Db 56 FLVSTIDANLADPREFLIILGSPKLTARGLSAPFLREGGKTDFLI-----DPKKEST 111
QY 137 GPGPDYLLKNYEDDVRSDVALDKQKCKIAQ-HPDVMLELQREAAQMHVLLKKEQF-- 193
Db 112 FEERSYQSOVNQDI-----CRYGSIPDPVEEKLREWPYQSQL-LLREHYQK 158
QY 194 ---SNYS-----NLIL----- 202
Db 159 KFNKSTYSRSSVDLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLDLYCSSKGYNI 218
QY 203 -----TEPNYRTHMHGRAVNSQLGKDYIQLKSLQPIRIYSRASLYGNIGRPRKNVIA 257
Db 219 SWELGNEPNSFKKADIFINGSLGEDFIQLHKLRR-STFNKAKLYGPDVGPQRKKTAK 277
QY 258 LLDGFMKAGSTVDATWQHCHYDGRVVKVMDFLKTRLDLTSLDQIRKIQKVNNTYTPGK 317
Db 278 MLKSLKAGGEVIDSVTWHYLYNGRATREDFLNPDVLDIFSSQKVFQVVESTPC 337
QY 318 KWLGEVWTTSGGTTNNLSDSYAAGFLMNTLGLMANQIDVYIRHSFFDHGNYHLVDQ 377

Db 338 KWLGETSSAYGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDEN 397
QY 378 ENPLPDYWLSLLYKRLIGPKVLAVHVGLOKRPGRVIRDKLRIYAHCTNHHNNHYVRG 437
Db 398 FDPPLDYWLSLLFKLVGTKVLMAVSGSKRR-----KLRVYLHCTNTDNPRIYK 448
QY 438 STTLFTINLHRSRKKIKLAGTLRDLKHQYLLQPYGOEGLKSKSVQNLGQPLVMVDDGTL 497
Db 449 DLTVAINLHNVTYKLRPLPYPSNKQVDKYLRLPLGPHGLLSKSVQNLGLTLKMWDDQTL 508
QY 498 PELKPRPLRAGRTL 511
Db 509 PPLMEKPLRPGSSL 522

RESULT 6
US-09-181-336-19
; Sequence 19, Application US/09181336A
; Patent No. 6242238
; GENERAL INFORMATION:
; APPLICANT: FREEMAN, Craig Geoffrey
; APPLICANT: PARISH, Richard
; APPLICANT: HAMDORF, Brenton James
; APPLICANT: HULETT, Mark Darren
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN
; FILE REFERENCE: 032505-004
; CURRENT APPLICATION NUMBER: US/09/181,336A
; CURRENT FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: AU PP0062
; EARLIER FILING DATE: 1997-10-28
; EARLIER APPLICATION NUMBER: AU PP0812
; EARLIER FILING DATE: 1997-12-09
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Rattus sp.
US-09-181-336-19

Query Match 28.9%; Score 805; DB 4; Length 380;
Best Local Similarity 47.3%; Pred. No. 1.9e-77;
Matches 155; Conservative 57; Mismatches 106; Indels 10; Gaps 2;

QY 204 EPNNYRTHMHGRAVNSQLGKDYIQLKSLQPIRIYSRASLYGNIGRPRKNVIALLDGPM 263
Db 62 EPNSFWKKAHISIDGLQGEDEVELHKLQK-SAFQNAKLYGPDIGQPKGTVKLLRSE 120
QY 264 KVAGSTVDATWQHCHYDGRVVKVMDFLKTRLDLTSLDQIRKIQKVNNTYTPGKKIWL 323
Db 121 KAGGEVIDSLTWHYLYNGRVATKEDFLSSDVLDTFILSVQILKVTKEMTGKKVWLGE 180
QY 324 VYTTGAGTNNLSDSYAAGFLMNTLGLMANQIDVYIRHSFFDHGNYHLVDQNFNPLPD 383
Db 181 TSSAYGGAPLLSDTFAAGFMWLDKLGSAQLGIEVVMRQVFFGAGNYHLVDENFEPLD 240
QY 384 YWLSLLYKRLIGPKVLAVHVGLOKRPGRVIRDKLRIYAHCTNHHNNHYVRGTTLFI 443
Db 241 YWLSLLFKLVGPVKLMSRVKGP-----RSKRLVYLHCTNHYHPRYREGDTLV 291
QY 444 INLHRSRKKIKLAGTLRDLKHQYLLQPYGOEGLKSKSVQNLGQPLVMVDDGTLPELKPR 503
Db 292 LNLHNVTYKLRPLPMPFSPVDKYLKPGSDGLLSKSVQNLGQTLKMWDEQTLPALTEK 351
QY 504 PLRAGRTLVPVMTGFFVVKVNNALAC 531
Db 352 PLPAGSSLSVPAPFSYGFVIRNAKTAAC 379

RESULT 7
US-09-181-336-17


```
; Sequence 17, Application US/09181336A
; Patent No. 6242238
; GENERAL INFORMATION:
; APPLICANT: FREEMAN, Craig Geoffrey
; APPLICANT: PARISH, Richard
; APPLICANT: HAMDORE, Brenton James
; APPLICANT: HULETT, Mark Darren
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN
; FILE OF INVENTION: ENDOLUCURONIDASE AND USES THEREFOR
; FILE REFERENCE: 032505-004
; CURRENT APPLICATION NUMBER: US/09/181,336A
; CURRENT FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: AU PP0062
; EARLIER FILING DATE: 1997-10-28
; EARLIER APPLICATION NUMBER: AU PP0812
; EARLIER FILING DATE: 1997-12-09
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Mus musculus
; US-09-181-336-17

Query Match      28.0%; Score 779; DB 4; Length 380;
Best Local Similarity 46.3%; Pred. No. 1.2e-74;
Matches 152; Conservative 56; Mismatches 110; Indels 10; Gaps 2;

QY 204 EPNVYTMHGRAVNGSOLGCKDYIOLKSLLOPIRIYSRASIYGNIGRPKNVIALDGGW 263
DB 62 EPNFVKKAHILDLGQLGDEFEVLHKLQR -SAFONAKLYGDIGDQPRGKTVKLLRSL 120

QY 264 KVAGSTVDATVQHCYIDGRVYKVMDFLTKRLDLSQIRKIQKVVNTYTPGKKIWLGE 323
DB 121 KAGGEVDSLTWHYLYNGRIATKEDFLSDVLDLTILSVOKILKYTKETPGKKVWLGE 180

QY 324 VVTSAGGTNNLSDSYAAGFLWNLITGLMANGIDVIRHSFFDHGYNHLVDQNFNPLPD 383
DB 181 TSYAGGGAPLNSNTFAAGFMFLDKLGLSAQMGIEVVRQVFGAGNYHLVDENFEPLPD 240

QY 384 YWLSLLYKRLIGPKVLAVHAGLQRPGRVIRDKRLRIYAHCTNNHNNHYRGSTLFI 443
DB 241 YWLSLLFKLVGPRVLLSRVKGPD-----RSKRLVYLHCTNYHPRYQEGDITLYV 291

QY 444 INLHRSRKKIKLAGTLRDLKLVHQLQPYQOGLKSKSVQLNGOPLVMVDDGTLPELKPR 503
DB 292 LNLHNVTKHLKVPPLFRKPVDTYLLKPSGPDGLLSKSVQLNGOILKMVDEQTLPALTEK 351

QY 504 PLRAGRTIVPPVTMGFFVVKVNNALAC 531
DB 352 PLPAGSALSLPAFSYGFYVIRDAKIAAC 379

RESULT 8
US-08-878-989-1
; Sequence 1, Application US/08878989
; Patent No. 5885803
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl G.
; APPLICANT: Lal, Preeti
; APPLICANT: Goli, Surya K.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESS: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
```

```
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/878,989
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0321 US
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 685 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HUVEBOB01
; CLONE: 39043
; US-08-878-989-1

Query Match      3.4%; Score 94.5; DB 2; Length 685;
Best Local Similarity 23.1%; Pred. No. 0.56;
Matches 84; Conservative 46; Mismatches 144; Indels 89; Gaps 18;

QY 129 RNPASRGSGPGPDYLLKNYEDDIVRSDVALDKQCKIAQHDPVMLELOREKAAQ----- 183
DB 84 RGVVLKGG-----FAKCYE-----MTDLTNKVAAKIIPHSRVAKPHQREKIDKEIELH 134

QY 184 -----MHLVLLKQFNTYSNLITPEPNVYTMHGRAVNGSOLGKDYIOLKSLLOP-IRI 237
DB 135 RILHHKVVQFYHYFEDKENIYILLEVCSSRM-----AHILK---ARKVLTEPEVRY 184

QY 238 YSRASLYGPNIGPRKNVIALDGFPMKVAGSTVDATVQHCYIDGRVYKVMDFLTKRLTD 297
DB 185 YLRQIVSGLKYLHEQE-----ILHRDLKLGNNFIN-----EAMELKVGDFGLAARLE 231

QY 298 TLDQIRKIKQVVNTYTP-----GKKIWLEGVV---TTSAG-----CTNNLSDSY-- 339
DB 232 PLEHRRRTICGTNYLSPEVLNKGHCESDIWALGCVMTMLLGRPPFTTNLKETYRC 291

QY 340 --AAGFLWLNTL-----GMLANQ-----GIDVYIRHSFFDHGY--NHLVDQNFNPL 381
DB 292 IREARYTMPSSLLAPAKHLIASMLSKNPEDRPSLDLDIRHDFLQGTPTDRLSSSCCHTV 351

QY 382 PDYWLSLLYKRLIGPKVLAVHAGLQRPKPR-----PGRVIRDKRLRIYAHCTNNHNNHYVRG 437
DB 352 PDFHLSPAKNFF-KKAAALFGGKKDKARYIDTHNRVSKEDIY-----KLRHDLKKT 405

QY 438 SIT 440
DB 406 SIT 408

RESULT 9
US-09-136-2-2
; Sequence 2, Application US/09136282
; Patent No. 6063609
; GENERAL INFORMATION:
```

APPLICANT: ANDERSON, KAREN
APPLICANT: JACKSON, JEFFREY
APPLICANT: HANSBURY, MICHAEL
APPLICANT: NERURKAR, SANDHYA
APPLICANT: ROSE, MARK
APPLICANT: BOUZYK, AMY
TITLE OF INVENTION: HUMAN SERUM INDUCIBLE KINASE (SNK)
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ratner & Prestia
STREET: P.O. Box 980
CITY: Valley Forge
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,282
FILING DATE: 20-AUG-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/056,112
FILING DATE: 20-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Prestia, Paul F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GH-70231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0700
TELEX: 846169
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 685 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-136-282-2

Query Match 3.4%; Score 94.5; DB 3; Length 685;
Best Local Similarity 23.1%; Pred. No. 0.56;
Matches 84; Conservative 46; Mismatches 144; Indels 89; Gaps 18;
QY 129 RNPASRGPGPDYILKNYEDDIVRSDVALDKQCKIAQHDPVDMLELQREKAAQ-----183
DB 84 RGVILKGG-----FACYE-----MTDLTNKVYAAKIIHPSRVAKPHOREKIDKELELH 134
QY 184 -----MHLVLLKEQFSNTYSLNLTPEPNYRTVTHGRAVNSQLGKDYIQLKSLQP-IRI 237
DB 135 RILHKKHVQVQFYHFKEDKENIYILLEYCSRSM-----AHILK-----ARKVLTEPEVRY 184
QY 238 YSRASLYGPNIGRPNKRVNIALDGFVKVAGSTVDVAVTQHCYIDGRVVKYMDFLKTRLLD 297
DB 185 YLRQIVSGLKYLHEQE-----ILHRDLKLGNNFFN-----EMELKVGDFGLAARLE 231
QY 298 TLDQIRKIQKRVNNTYTP-----GKKIWLEGVY--TTSAG-----GTNNLSDSY-- 339
DB 232 PLEHRRRTICGTPNYSPEVLNKGHCESDIWALGCVMTLLGRPPFTTNLKETRYC 291
QY 340 --AAGFLWNL-----GMLAQ-----GIDVVIRISFDHGY--NHLVDQNFPL 381
DB 292 IREARYTWPSSLLAPAKHLIASMLSKNPEDRPSLDDIIRHDFLQGFDPRLSSSCCHY 351
QY 382 PDVWLSLLYKRLIGPKVLAVHAGLQKRP-----PGRVIRDKLRIYAHCTNHNHNHNVVRG 437
DB 352 PDFHLSFPKNFF-KKAAALFGGKKDKARYIDTHNRVSKEDDIY-----KLHRDLKKT 405

QY 438 SIT 440
DB 406 SIT 408
RESULT 10
US-09-272-796-1
; Sequence 1, Application US/09272796
; Patent No. 6207148
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl G.
; APPLICANT: Lal, Preeti
; APPLICANT: Goli, Suriya K.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN
; TITLE OF INVENTION: KINASES
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/272,796
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/878,989
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0321 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 685 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HUVENOB01
; CLONE: 39043
; US-09-272-796-1

Query Match 3.4%; Score 94.5; DB 4; Length 685;
Best Local Similarity 23.1%; Pred. No. 0.56;
Matches 84; Conservative 46; Mismatches 144; Indels 89; Gaps 18;
QY 129 RNPASRGPGPDYILKNYEDDIVRSDVALDKQCKIAQHDPVDMLELQREKAAQ-----183
DB 84 RGVILKGG-----FACYE-----MTDLTNKVYAAKIIHPSRVAKPHOREKIDKELELH 134
QY 184 -----MHLVLLKEQFSNTYSLNLTPEPNYRTVTHGRAVNSQLGKDYIQLKSLQP-IRI 237
DB 135 RILHKKHVQVQFYHFKEDKENIYILLEYCSRSM-----AHILK-----ARKVLTEPEVRY 184
QY 238 YSRASLYGPNIGRPNKRVNIALDGFVKVAGSTVDVAVTQHCYIDGRVVKYMDFLKTRLLD 297

Db 190 DTFSQISQGLTKLVNKGGLTGLDKLNKFSGSKAGLGLVVISGLLSGATAALVLADKN 249

QY 340 A-----AGFLWNLTLGMLANQ 355

Db 250 ASTDRKVGAGF-----ELANQ 265

RESULT 13

US-08-257-341-9

; Sequence 9, Application US/08257341

; Patent No. 5525491

; GENERAL INFORMATION:

; APPLICANT: HUSTON, JAMES S

; APPLICANT: OPPERMANN, HERMANN

; APPLICANT: TIMASHEFF, SERGE N

; TITLE OF INVENTION: SERINE RICH PEPTIDE LINKER

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CREATIVE BIOMOLECULES, INC./PATENT DEPT.

; STREET: 35 SOUTH STREET

; CITY: HOPKINTON

; STATE: MA

; COUNTRY: USA

; ZIP: 01748

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/257,341

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION NUMBER: US/07/842,149

; APPLICATION DATA:

; FILING DATE:

; APPLICATION NUMBER: US 07/662,226

; FILING DATE: 27-FEB-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: CAMPBELL ESQ, PAULA A

; REGISTRATION NUMBER: 32,503

; REFERENCE/DOCKET NUMBER: CRP-064CP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617/248-7000 (ATTY)

; INFORMATION FOR SEQ ID NO: 9:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 471 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-257-341-9

Query Match 3.3%; Score 93; DB 1; Length 471;

Best Local Similarity 19.2%; Pred. No. 0.43;

Matches 90; Conservative 71; Mismatches 152; Indels 156; Gaps 21;

QY 55 LKKTLLDVSTKNPVRTVNEFLSLQDPSIIHGWLDLSSKRLVTLARGLS----- 109

Db 68 LKDRINIVLSRELKEPR--GAHFLAKSLDRLRLQPELASKVDVMVIGSSVYQEA 125

QY 110 ---PAFLRFGGKRTDFLQFQNLNRPASRGPGDYLYLKNYEDIVRSVDALDKRGCKI 166

Db 126 MNQPGHLRLVTR-----IMQEFESDTFFPEIDLKGYK--LL 160

QY 167 AQHPDVMLELQREKAAQMHVLVLLKEQFSNTYSNLIILTEPNNYTMGRAVNSQLGKDYI 226

Db 161 PEYGVLSVEOEKGIKFEVYEKKDAS-----VTVSSSGSSSSSGSSSMIEOD-- 212

QY 227 QLSKLLQPIRIYSRASLYGNIPRKNRIALL--DGFMKVGASTVDVATWQHCHYIDGRV 284

Db 213 -----GLHAGSPAANVVERLFGYDWAQQTICGS--DAAVFR--LSAGGRP 252

QY 285 VKVMDFLKRLDLDTSQIRKIQKVVNTYTPGKKIWLEGVVTTSGAGTNNLSDSYAAGFL 344

Db 253 VL---FVKTDL-----SGALNELQDE--AARLS 275

QY 345 WLNTGLMLANOGIDVIRHSHFFDGHGYNHLVDQNFNPLPDYWLSLLYKRLIGPKVLAVHVA 404

Db 276 WLATTGVPCAAVLDVVT-----EAGRD-----W--LLLGEVPGQDLSLHLA 315

QY 405 GLQKRPGRVIRDKLRIY-----AHCTNHHNHNVVRGSIITLFIINLHRSRKKIKLAGTL 459

Db 316 PAEKY-----SIMADAMRLHLTDPATCFDHOAKH-----RIERATRWAEAGLYD 361

QY 460 RDKLVHVOYLLQPYGOEGL-----KSKSVQLNGQPLVMV--DDGTLPEL 500

Db 362 QDDLDEH-----OGLAPAEFLARLKARMPDGEDLVVTHGDACLPMI 403

RESULT 14

US-09-346-237-2

; Sequence 2, Application US/09346237A

; Patent No. 6265197

; GENERAL INFORMATION:

; APPLICANT: Bisgard-Frantzen, Henrik

; APPLICANT: Svendsen, Allan

; TITLE OF INVENTION: Starch Debranching Enzymes

; FILE REFERENCE: 5629.200-US

; CURRENT APPLICATION NUMBER: US/09/346.237A

; CURRENT FILING DATE: 1999-07-01

; EARLIER APPLICATION NUMBER: PA 1998 00868

; EARLIER FILING DATE: 1998-07-02

; EARLIER APPLICATION NUMBER: 60/094,353

; EARLIER FILING DATE: 1998-07-28

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 915

; TYPE: PNT

; ORGANISM: Bacillus deramificans

; FEATURE:

; NAME/KEY: PEPTIDE

; LOCATION: (1)-(915)

; OTHER INFORMATION: Pullulanase

US-09-346-237-2

Query Match 3.3%; Score 91.5; DB 4; Length 915;

Best Local Similarity 23.0%; Pred. No. 1.9;

Matches 79; Conservative 38; Mismatches 127; Indels 99; Gaps 19;

QY 43 DRRLPVDRAAGLKE--KTLILLDVSTKNP--VRTVNEFLSLQDPSIIHGWLDLFLSS 98

Db 424 DVRDFSIDPNSGMKNKGYLALTEKGTGPDNVKT-----GID--SL 463

QY 99 KRLVTLARGLSPAFLRFGGKRTDFLQFQNLNRPASRGPGDYLYLKNYEDDIVRSVDAL 158

Db 464 KQLGTHVOLMVFASNSVDETD-----PTDNMGYDP-----RNY--DVPEGQYAT 508

QY 159 DKQKCKIAQHPDVMLELQREKAAQMHVLVLLKEQFSNTYSNLIILTEPN--NYRTHM--GRAV 216

Db 509 NANGNARIKEFEKREMLVSLHREHIGVMVYVNHTEFATQISDFDKIVPEYYRTDDAGNYT 568

QY 217 NGSQGLKDYIQLKSLLOPIRIYSRASLYGNIPRKNRIALLDGMFKVGASTVDVATW 275

Db 569 NGSGTGNETAERPMVOKF-----IIDSCLKYM 595

QY 276 QHCY-IDGRVVKVMDPLKTRLLDTSQIRKIQKVVNTYTPGKKIWLEGVVTTSGAGTNN 334

Db 596 VNEYHIDGRFDMALLGK---DTMS-----KASELHAINPG--ITALYCEPWT--GGTSA 644

QY 335 LSDSYAAGFLWNLTLGMLANQIDVIRHSHFFDGHGYNHLVDQN 377

Db 645 LPDDQ-----LLTKG--AOKGMGAV-----FNDNLRLNALDGN 675

Search completed: July 30, 2002, 08:15:01
Job time: 283 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:08:23 ; Search time 141.47 Seconds
(without alignments)
419.265 Million cell updates/sec

Title: US-09-836-461-2
Perfect score: 2785
Sequence: 1 MRVLCAPPEAMPSSNSRPPA.....PVTMGFFVKNVNALACRYR 534

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq_032802.*			
1:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*		
2:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*		
3:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*		
4:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*		
5:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*		
6:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*		
7:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*		
8:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*		
9:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*		
10:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*		
11:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*		
12:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*		
13:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*		
14:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*		
15:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*		
16:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*		
17:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*		
18:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*		
19:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*		
20:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*		
21:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*		
22:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2785	100.0	534	22	AAB85216
2	2746	98.6	592	22	AAU07424
3	2746	98.6	592	22	AAB85215
4	2746	98.6	592	22	AAU07632
5	2736	98.2	592	22	AAB81062
6	2564	92.1	492	22	AAB84664
7	2485	89.2	538	22	AAU07633
8	2478	89.0	480	22	AAU07418
9	2478	89.0	480	22	AAB85217
10	2478	89.0	480	22	AAU07634
11	2270	81.5	439	22	AAU07423

12	1314	47.2	262	22	AAM24147	Human EST encoded
13	1112	39.9	214	22	AAU99905	Human excretory re
14	1112	39.9	214	22	AAU43704	Human bladder anti
15	944.5	33.9	535	21	AAB08851	A murine heparanas
16	936.5	33.6	543	20	AAU02345	A human heparanas
17	936.5	33.6	543	21	AAB08849	Amino acid sequenc
18	936.5	33.6	543	21	AAU57590	Human heparanas
19	936.5	33.6	543	21	AAU52990	Human heparanas
20	936.5	33.6	543	22	AAU97635	Human heparanas
21	936.5	33.6	592	20	AAU02346	A human heparanas
22	936.5	33.6	592	21	AAB08850	Amino acid sequenc
23	933.5	33.5	543	22	AAB88361	Human membrane or
24	932.5	33.5	543	20	AAU17082	Human heparanas
25	932.5	33.5	543	22	AAB86206	Human heparanas
26	932.5	33.5	588	20	AAU30124	A human protein wi
27	929.5	33.4	530	20	AAU34173	Human pre-prohepar
28	889.5	31.9	532	20	AAU17083	Seq ID No: 15 Of W
29	805	28.9	380	20	AAU17085	Rat heparanas enz
30	779	28.0	380	20	AAU17084	Mouse heparanas
31	778	27.9	156	22	AAG65963	Human heparanas-1
32	468.5	16.8	488	22	AAB31469	Human heparanas-1
33	440.5	15.8	488	22	AAB31472	Amino acid sequenc
34	439.5	15.8	488	22	AAB31470	Amino acid sequenc
35	422.5	15.2	488	22	AAB31471	Amino acid sequenc
36	278	10.0	137	22	AAG65961	Human heparanas-1
37	278	10.0	159	22	AAG65964	Human heparanas-1
38	270	9.7	112	22	AAU07425	Human heparanas-1
39	247	8.9	256	21	AAG13479	Arabidopsis thalia
40	203	7.3	118	22	AAG65962	Human heparanas-1
41	171	6.1	935	22	AAB69219	Drosophila melanog
42	162	5.8	115	22	AAB85218	Mouse heparanas-1
43	144	5.2	38	20	AAU34186	Human pre-prohepar
44	127	4.6	24	22	AAB85220	Heparanas-like pr
45	127	4.6	25	22	AAB85221	Heparanas-like pr

ALIGNMENTS

RESULT 1
AAB85216
ID AAB85216 standard; Protein; 534 AA.

XX	AC	AAB85216;
XX	AC	
XX	DT	07-SEP-2001 (first entry)
XX	DT	
XX	DE	Heparanas-like protein Hpa2 splice variant #2.
XX	DE	
XX	KW	Heparanas; splice variant; homologue; heparanas-like protein; Hpa2;
XX	KW	cytostatic; neuroprotective; cerebroprotective; immunosuppressive;
XX	KW	antipsoriatic; nootropic; antiinflammatory; antiarthritic; antiasthmatic;
XX	KW	antidiabetic; antiarteriosclerotic; vulnerary.
XX	OS	Homo sapiens.
XX	PN	WO200146392-A2.
XX	PD	28-JUN-2001.
XX	PF	21-DEC-2000; 2000WO-GB04963.
XX	PR	22-DEC-1999; 99GB-0030392.
XX	PR	07-APR-2000; 2000GB-0008713.
XX	PA	(OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX	PI	McKenzie EA, Stamps AC, Terrett JA, Tyson KL;
XX	DR	WPI: 2001-418056/44.
XX	DR	N-PSDB; AAB22672.
XX	PT	Novel homologs of heparanas, present in three splice variants, useful

→ post dated

for identifying agents that modulate heparanase, useful in the treatment and/or prophylaxis of abnormal levels of heparanase

Claim 1: Fig 2; 97pp; English.

The invention provides a homologue to heparanase which is present in three splice variants. The heparanase homologue polypeptide is useful in the treatment of a human or non-human animal or for use in diagnosis. Vectors comprising the heparanase homologue polynucleotides are useful in the transformation or transfection of a prokaryotic or eukaryotic host. The modulators of the polypeptide are useful in the manufacture of a medicament for the treatment and/or prophylaxis of a condition/disease associated with abnormal levels of the heparanase homologue, including cancer, central nervous system (CNS) and neurodegenerative diseases, cardiovascular diseases such as stenosis following angioplasty and atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus, allografts, inflammatory diseases, arthritis, vascular stenosis, tumour growth and progression, asthma, Alzheimer's disease, diabetic retinopathy, wound healing and inflammation. The polypeptide is also useful in diagnosis and research. The present sequence represents the amino acid sequence of the mid-sized splice variant of the heparanase-like protein Hpa2 of the invention.

Sequence 534 AA;

```
Query Match      100.0%; Score 2785; DB 22; Length 534;
Best Local Similarity 100.0%; Pred. No. 8.3e-274;
Matches 534; Conservative 0; Mismatches 0; Indels 0;
```

1	MrVLCaPeAMPSNSRRPaCLaPaGALYLaLLHLuSSoAGDRRPuVpDRAAGLKEkTL	60
Qy		
1	MrVLCaPeAMPSNSRRPaCLaPaGALYLaLLHLuSSoAGDRRPuVpDRAAGLKEkTL	60
Db		
61	ILLDVSTKNPVRTNENfLSQLDPSIIHDGWLDFSSKRLVTLaRGLSPaflRFGGKT	120
Qy		
61	ILLDVSTKNPVRTNENfLSQLDPSIIHDGWLDFSSKRLVTLaRGLSPaflRFGGKT	120
Db		
121	DLQFQNLrNPakSRGGPGpDYLYkNYEDDIvRSDoVALDKOKGCKIAQHDPVMLEQREK	180
Qy		
121	dflqfqnLrnpakSRGGPGpDYLYknyeddiVrsdvaLdkgkckiaqhdpvmleIgrek	180
Db		
181	AAQHVLlLKbQFSNTYsNLILTEPNNYRTMHGRAVNSQLGKDYIOLKSLILOPIRYSR	240
Qy		
181	aaqhvlvlLKbqfsntysnLilTEpnnYrtmhgravnsglqgkdyIolkSLILOPIrYsR	240
Db		
241	ASLYGPNIGRPKNKIALLDGFMKvAGSTVDaVTHQHCYIDGRVVKVMDFLKTRLLDTLS	300
Qy		
241	aslygpnIgrknvIALldgfmkVagstvdavTHqhcYIdGRvVKVMDfLkTRllDtLS	300
Db		
301	DOIRIKQVWVTYTPGKKIWLEGVVTTSAGGTTNLSDSYAAGfLWLNLTGLMANOGIDVV	360
Qy		
301	dGIRIKqVwVtYtPGkKIwLEGvvtTsaggTtnLSdsYAAGfLwLntGLmAnoGIdVv	360
Db		
361	IRHSFFDHGYNHLVDQNFNLPDYWLSSLYIKRLIGPKVLAVHVAGLQRPGRVIRDKL	420
Qy		
361	IrHsffDHgYnHlVdQnFNlPDyWlSSlYIKrLIgPKvLAvHVagLqRPgRvIrDkL	420
Db		
421	RIYAChTNNHNNHYVRGSITFLINLHRSEKKIKLAGLIROKLHYOYLLOPYGOEGLKSR	480
Qy		
421	rIYAchTnHnHnHYvRGSITfLiNlHrSEKKlAGLIRokLHYoYLLOpYGOeGLKsR	480
Db		
481	SVOLMGQPLVMVDDGTLPELKPRLPAGRTLPIPPVTMGFFVVKNNVALACRYR	534
Qy		
481	svolmgQPLVMvDDGTLpELKpRlPaGRTLPIppVTMGffVvKNNvAlAcRYR	534
Db		

RESULT 2
AAU07424
IP: AAU07424 standard: Protein: 592 AA:

AAU07424;

18-DEC-2001 (first entry)

Human heparanase-like protein splice variant #1.

Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic; antiproliferative; cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial; virucide; fungicide; ophthalmological; extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis; hyperproliferative disorder; neoplasia; cardiovascular disorder; cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection; nervous system disorder; Alzheimer's disease; ocular disorder; sunburn; wound healing; food additive; heparanase.

Homo sapiens.

WO200179253-A1.

25-OCT-2001.

11-APR-2001; 2001WO-US11643.

18-APR-2000; 2000US-198123P.

(HUMA-) HUMAN GENOME SCI INC.

Fiscella M, Shi Y, Ebner R, Ruben SM;

WPI: 2001-611720/70.

N-PSDB: AAS13848.

New nucleic acids encoding extracellular matrix polypeptides, for diagnosing, treating, preventing or ameliorating human disorders and disease, such as, autoimmune, hyperproliferative or cardiovascular disorders -

Disclosure; Page 14; 308pp; English.

The invention relates to novel isolated polynucleotides (I) encoding extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by (I) are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. The antibodies to the polypeptides can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunoassay assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angioneurosis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. The present sequence represents the amino acid sequence of human heparanase-like protein, splice variant #1.

Sequence 592 AA;

Query Match 98.6%; Score 2746; DB 22; Length 592;

Best Local Similarity 90.2%; Pred. No. 9.1e-270;

Matches 534; Conservative 0; Mismatches 0; Indels 58; Gaps 1

Db	61	illdvstknprtvenflslqlpsihdglwldflslskrlvtlrglspafirfggkrt	120
Qy	121	DFLQFQNLNRPKSRGGPGDYLYLKNYEDDIVRSDVALDKQKCKIAQHPDVMLELQREK	180
Db	121	dflqfqnlnrpakargpgpdylylknayeddivrsdvaldkqkckiaqhpdmlelqrek	180
Qy	181	AAQMHVLVLKEQFSNTYSNLIIT-----	203
Db	181	aaqmhlvlkqefsnltysnliltarsldklynfadcsghlilfalnlnrrpnnswnss	240
Qy	204	-----EPNNYRTMHGRAVNGSOLGKDYIQLKSLLOPIRIYSRAS	242
Db	241	alsllkysaskyniswelgnepnnyrtmhgravnsglqkdyiqkllqpiriysras	300
Qy	243	LYGNIGRPRKNVIALLDGFMKVAGSTVDAYTWQHICYIDGRVVKVMDFLKRLDLSQ	302
Db	301	lygpnigrprknvialldgfmkvagstvdavtwqhcycidgrvvkvmdfiktrlldtsdq	360
Qy	303	IRKIQKVVNTYTPGKKIWLQGVVTSAGTNNLSDSYAAGFLWNLGMLANQGDVIVIR	362
Db	361	irkigkvvntytpgkkiwlegvvttsagtnnlsdsyaagflwntlgmlanqgidvivr	420
Qy	363	HSFFDHGYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHVAGLQKRPGRVIRDKLRI	422
Db	421	hsffdhgynhlvdqnfnpldywlssllykrligpkvlavhvaglqkripgrvirdklri	480
Qy	423	YAHCTNNHHNHYVRSITFLIINLHRSRKKIKLAGTLRDLKHVHOYLQPYGQGLKSKSV	482
Db	481	yahctnnhnhnyvrsitfliflnhrrsrkkiklagtlrdklvhqylqpygqeglksksv	540
Qy	483	QLNGQPLVMVDDGTLPELKPRPLRAGRPLVIPPVTMGFFVVKVNNALACRYR	534
Db	541	qlngqplvmvddgtlpelkprplragrtlivppvtmgffvkvnnvalacryr	592
RESULT	3		
AAB85215			
ID	AAB85215	standard; Protein; 592 AA.	
XX			
AC	AAB85215;		
XX			
DT	07-SEP-2001	(first entry)	
XX			
DE	Heparanase-like protein Hpa2 splice variant #1.		
XX			
KW	Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;		
KW	cytostatic; neuroprotective; cerebroprotective; immunosuppressive;		
KW	antipsoriatic; nontropic; antiinflammatory; antiarthritic; antiasthmatic;		
KW	antidiabetic; antiarteriosclerotic; vulnerary.		
XX			
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	Misc-difference 237	/label= unknown	
FT		/note= "encoded by ANC"	
FT			
XX	WO200146392-A2.		
XX			
PD	28-JUN-2001.		
XX			
PF	21-DEC-2000; 2000WO-GB04963.		
XX			
XX	22-DEC-1999; 99GB-0030392.		
PR	07-APR-2000; 2000GB-0008713.		
XX			
PA	(OXFO-) OXFORD GLYCOSCIENCES UK LTD.		
XX			
PI	McKenzie EA, Stamps AC, Terrett JA, Tyson KL;		
XX			
DR	WPI; 2001-418056/44.		
DR	N-PSDB; AAB22671.		
XX			

PT	Novel homologs of heparanase, present in three splice variants, useful for identifying agents that modulate heparanase, useful in the treatment and/or prophylaxis of abnormal levels of heparanase -		
PT	Claim 1; Fig 1; 97pp; English.		
XX			
CC	The invention provides a homologue to heparanase which is present in three splice variants. The heparanase homologue polypeptide is useful in the treatment of a human or non-human animal or for use in diagnosis.		
CC	Vectors comprising the heparanase homologue polynucleotides are useful in the transformation or transfection of a prokaryotic or eukaryotic host.		
CC	The modulators of the polypeptide are useful in the manufacture of a medicament for the treatment and/or prophylaxis of a condition/disease associated with abnormal levels of the heparanase homologue, including cancer, central nervous system (CNS) and neurodegenerative diseases, cardiovascular diseases such as stenosis following angioplasty and atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus, allografts, inflammatory diseases, arthritis, vascular restenosis, tumour growth and progression, asthma, Alzheimer's disease, diabetic retinopathy, wound healing and inflammation. The polypeptide is also useful in diagnosis and research. The present sequence represents the amino acid sequence of the largest splice variant of the heparanase-like protein Hpa2 of the invention.		
XX	Sequence 592 AA;		
SQ			
	Query Match	98.6%; Score 2746; DB 22: Length 592;	
	Best Local Similarity	90.2%; Pred No. 9,1e-270;	
	Matches 534; Conservative 0; Mismatches 0; Indels 58; Gaps 1;		
Qy	1	MRVLCAFPEAMPSSNRPPACIAPGALYLAALLHLSSLSSQAGDRRLPVDRAAGLKERTL	60
Db	1	mrvlcafpeampssnrppaclapgalyallhlhslssqagdrprlpvdraaglkertl	60
Qy	61	ILLDVSTKNPVTNENFLSLQDPSIHDGWLDFLSSKRLVTLARGLSPAFLRGGKRT	120
Db	61	illdvstknprtvenflslqlpsihdglwldflslskrlvtlrglspafirfggkrt	120
Qy	121	DFLQFQNLNRPKSRGGPGDYLYLKNYEDDIVRSDVALDKQKCKIAQHPDVMLELQREK	180
Db	121	dflqfqnlnrpakargpgpdylylknayeddivrsdvaldkqkckiaqhpdmlelqrek	180
Qy	181	AAQMHVLVLKEQFSNTYSNLIIT-----	203
Db	181	aaqmhlvlkqefsnltysnliltarsldklynfadcsghlilfalnlnrrpnnswnss	240
Qy	204	-----EPNNYRTMHGRAVNGSOLGKDYIQLKSLLOPIRIYSRAS	242
Db	241	alsllkysaskyniswelgnepnnyrtmhgravnsglqkdyiqkllqpiriysras	300
Qy	243	LYGNIGRPRKNVIALLDGFMKVAGSTVDAYTWQHICYIDGRVVKVMDFLKRLDLSQ	302
Db	301	lygpnigrprknvialldgfmkvagstvdavtwqhcycidgrvvkvmdfiktrlldtsdq	360
Qy	303	IRKIQKVVNTYTPGKKIWLQGVVTSAGTNNLSDSYAAGFLWNLGMLANQGDVIVIR	362
Db	361	irkigkvvntytpgkkiwlegvvttsagtnnlsdsyaagflwntlgmlanqgidvivr	420
Qy	363	HSFFDHGYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHVAGLQKRPGRVIRDKLRI	422
Db	421	hsffdhgynhlvdqnfnpldywlssllykrligpkvlavhvaglqkripgrvirdklri	480
Qy	423	YAHCTNNHHNHYVRSITFLIINLHRSRKKIKLAGTLRDLKHVHOYLQPYGQGLKSKSV	482
Db	481	yahctnnhnhnyvrsitfliflnhrrsrkkiklagtlrdklvhqylqpygqeglksksv	540
Qy	483	QLNGQPLVMVDDGTLPELKPRPLRAGRPLVIPPVTMGFFVVKVNNALACRYR	534
Db	541	qlngqplvmvddgtlpelkprplragrtlivppvtmgffvkvnnvalacryr	592
RESULT	4		

CC associated with neoangiogenesis and metastasis related to malignant
CC tumour formation. Heparanase-2 polynucleotides and proteins are useful as
CC vaccines for inducing an immunological response against autoimmune
CC disorders, blood coagulation disorders, cancer, diabetes, ischaemia,
CC sepsis, stroke, cardiovascular diseases, or thrombosis, as well as in
CC diagnosing (the susceptibility of a subject to) these diseases
CC Heparanase-2 fragments may be used as immunogens to produce antibodies
CC immunospecific to the polypeptides, and to identify membrane bound
CC soluble receptors, agonists or antagonists that compete with the binding
CC of the polypeptide to the receptors. An antibody specific for
CC heparanase-2 can be used in the diagnosis of the above diseases and in
CC isolating or identifying clones expressing heparanase-2. The present
CC sequence represents heparanase-2. Three regions of heparanase-2 with high
CC immunogenicity (immunogenic epitopes) can be used to raise antibodies
CC against heparanase-2.
XX
SQ Sequence 592 AA;

Query Match 98.2%; Score 2736; DB 22; Length 592;
Best Local Similarity 89.9%; Pred. No. 9.4e-269;
Matches 532; Conservative 1; Mismatches 1; Indels 58; Gaps 1;

Qy 1 MRVLCAPPEAMPSSSRPACIAPGALYALALLHLSSQAGDRPLPVDRAAGLKEKTL 60
Db 1 mrvlcafpampsssrpaciapgalylalallhlslssqagdrplpvdraaglikektl 60
Qy 61 ILDDVSTKNPVRTVNEFSLQLDPSIIHDGWLDFSLSKRLVTLARGSPAFLRFGGKRT 120
Db 61 illdvstknprtvenfslqldpsiihdgwldfslskrlvtlargspaflrfggkrt 120
Qy 121 DFLQFONLRNPAKSGGPGDYLYKNYEDDIVRSDVALDKQKCKIAQHPDVMLELQREK 180
Db 121 dflqfqnlnpaksrggpgdylyknyeddivrsdvaldkqkckiaqhpdvmlvlgrek 180
Qy 181 AAQHMLVLLKQFSNTYNNLIIT----- 203
Db 181 aaqhmlvllkqfsntynnlit----- 203
Qy 204 -----EPNNYRTMHGRAVNGSOLGKDYIOLKSLLOPIRYSRAS 242
Db 241 alsllkysasknyslswelgnepnnyrtmhgravgsglqkdyiklsllqpirysras 300
Qy 243 LYGPNIQPRKNVIALDGFMKVAGSTVDVAVTWHQCYIDGRVVKVMDFLKRLDLTSLDQ 302
Db 301 lygpnigrprknvialldgfmkvagstvdavtwqhcycidgrvvkvmfdlkrldtldsq 360
Qy 303 IRKIQKVVNTYTPGKKIWLEGVVTTSAGTNNLSDSYAGFLWNTLGLANQGDIVVIR 362
Db 361 irkikqvntytpgkklwlegvvttsaggttnlssdsyagflwntlglanqgdvivr 420
Qy 363 HSFFDHGYNHLVDQNFPLPDYWLSLLYKRLTGPKVLAVHVGAGLQKPRGVRIRDKLRI 422
Db 421 hsfddhgynhlvdqnfplpdylwsllykrltgpkvlavhvgaglrkprgvrirdklri 480
Qy 423 YAHCTNNHNNHYVRSITFLIINLHRSRKKIKLAGTLRDLKVHQLQPYGQGLKRSV 482
Db 481 yahctnnhnnhyvrsitflinlhrsrrkkikagtlrldklvhqyllypgyqeglkrsv 540
Qy 483 QLNGQPLVMVDGDTLPELKPRLRAGRLVIPPVTMGFFVKNVNALACRYR 534
Db 541 qlngqplvmvdgdtlpeikprlragrtlvipvtmgffvknvnalacryr 592

RESULT 6
RAB84664
SD AAB84664 standard; Protein; 492 AA.

XX AAB84664;
XX AC
XX DT 05-SEP-2001 (first entry)
XX

DE Amino acid sequence of human heparanase-like polypeptide.

XX Human; heparanase-like polypeptide; gene therapy; cancer; angiogenesis;
KW trauma; autoimmune disease; skin disease; cardiovascular disease;
KW nervous system disease; inflammation; arthritis; genitalia;
KW male fertility; erectile dysfunction.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 407
FT /note= "unspecified residue encoded by KCA"
XX WO200148161-A2.
XX 05-JUL-2001.
XX
PF 18-DEC-2000; 2000WO-EP12909.
XX
PR 23-DEC-1999; 99EP-0125831.
XX
PA (SCHD) SCHERING AG.
XX
PI Siemeister G, Weiss B;
XX
DR WPI; 2001-418259/44.
DR N-PSDB; AAH28347.
XX
XX Human Heparanase-like polynucleotide encoding polypeptides useful for
PT modulating expression of the polypeptide and for treating cancer, -
PT cancer metastasis, aberrant angiogenesis by gene therapy technique -
XX
PS Claim 9; Page 30; 30pp; English.
XX
CC The present sequence represents a human heparanase-like polypeptide.
CC Heparanase-like polynucleotides are useful as a source of probes, like
CC primers and antisense molecules, and in gene therapy. Heparanase-like
CC polynucleotides and polypeptides are useful for treating several
CC disorders e.g., cancer, cancer metastasis. The oligonucleotides are
CC also useful as diagnostic markers for the diagnosis of disorder such
CC as cancer, cancer metastasis and aberrant angiogenesis. They may also
CC act as diagnostic markers for diagnosis of disorder such as cancer,
CC cancer metastasis and aberrant angiogenesis. The heparanase polypeptides
CC and polynucleotides are also useful for treating trauma, autoimmune
CC diseases, skin diseases, cardiovascular diseases, nervous system
CC diseases, and inflammation including arthritis. Since the polynucleotide
CC is preferentially expressed in male genitalia, modulation of its
CC expression and/or activity may be used for medical intervention in male
CC genitalia function that is male fertility control, erectile dysfunction.
XX
SQ Sequence 492 AA;

Query Match 92.1%; Score 2564; DB 22; Length 492;
Best Local Similarity 99.6%; Pred. No. 2.1e-251;
Matches 490; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 43 DRRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNEFSLQLDPSIIHDGWLDFLSKRLV 102
Db 1 drrplpvdraaglikektlilldvstknprtvenfslqlpsiihdgwldfllskrlv 60
Qy 103 TLARGSPAFLRFGGKRTDFLOFQNLNRPNAKSGGPGPDYLYKNYEDDIVRSDVALDKQK 162
Db 61 tlargspaflrfggkrtdflofqnlnrpaksggpgpdylyknyeddivrsdvaldkqk 120
Qy 163 GCKIAQHPDVMLELOBEKAAQHMLVLLKQFSNTYNNLSLITTEPNYRTMHGRAVNGSQLG 222
Db 121 gckiaqhpdvmllelgrekaagmhlvlkqfsntynliltepnnyrtmhgravgsqlg 180
Qy 223 KDVIOLKSLLOPIRYSRASLYGPNIGRPRKNVIALDGFMKVAGSTVDVAVTWHQCYIDG 282
Db 181 kdviolksllopiriysraslygpnigrprknvialldgfmkvagstvdavtwqhcycidg 240
Qy 283 RVVKVMDFLKRLDITLSDQIRKIQKVVNTYTPGKKIWLEGVVTTSAGTNNLSDSYAG 342

XX (HUMA-) HUMAN GENOME SCI INC.
XX Fiscella M, Shi Y, Ebner R, Ruben SM;
PI WPI; 2001-611720/70.
DR N-PSDB; AAS13843.
XX New nucleic acids encoding extracellular matrix polypeptides, for
PT diagnosing, treating, preventing or ameliorating human disorders and
PT disease, such as, autoimmune, hyperproliferative or cardiovascular
PT disorders -
XX Claim 1; Page 292-293; 308pp; English.
XX The invention relates to novel isolated polynucleotides (I) encoding
CC extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by
CC (I) are used to prevent, treat or ameliorate a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility to
CC a pathological condition. The antibodies to the polypeptides can also be
CC used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities. The present sequence represents the amino acid sequence
CC of novel human extracellular matrix (ECM) protein #1.
XX Sequence 480 AA;
SQ

Query Match 89.0%; Score 2478; DB 22; Length 480;
Best Local Similarity 89.9%; Pred. No. 1.1e-242;
Matches 480; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

QY 1 MRVLCAPPEAMPSSNRPPACIAPGALYLALLHLHLSSQAGDRPLPVDRAAGLKEKTL 60
DB 1 mrvlcafpampssnrppaciapgalylallhlhlssqagdrplpvdraaglkectl 60
QY 61 ILLDVSTKPNRTVNFSLQLDPSIIHDGWLDFLSSKRLVTLARGLSPAFLRFGGKRT 120
DB 61 illdvstknprtvnfnslqldpsiihdgwlfdflsskrlvltlrglspafllrfggkrt 120
QY 121 DFLQFONLRNPAKSRGPGPDYLYKNYEDDIVRSVDALDKQCKIAQHPDVMLELQREK 180
DB 121 dflqfonlrnpaksrpgpgpdylyknyeddivrsvdaldkqckiaqhpdvmlelqrek 149
QY 181 AQMHVLVLLKEQFSNTYSLNLIITEPNRYRTMHGRAVNSQLGRDYIQLKSLLPRIYSR 240
DB 150 -----epnyrtmhgravnsglqgdyqlksallqpiirysr 186
QY 241 ASLYGNIGRPNKNTALLDGFMKVAGSTVDATWQHCVIDGRVVKVWDFLKRLLDTLS 300
DB 187 aslygnigrpnkntalldgmfvagstvdavtwhcvidgrvrvkvmfklrlltdtls 246
QY 301 DQIRKIQKVNTYTPGKKIWLQGVVTTSGAGTNNLSDSYAAGFLWLTGLMLANOGIDVV 360
DB 247 dqirkikvntytpgkklwlqgvvttsaggtunlstdsyaagflwltgmlangidvv 306
QY 361 IRHSFFDHGYNHLVDQNFPLPDYWLSLLYKRLIGPKVLAVHVAGLQKRPGRVIRDKL 420
DB 307 irhsffdhgynhlvdqnfplpdywlslllykrligpkvlavhvaglqkrpgrvirdkl 366

QY 421 RIYAHCTNHHNHNVRGSTITLFIINLHRSRKKIKLAGTLRDLKHQVLLQOPYGQEGLSK 480
DB 367 riyahctnhnhnvrgstitlfiinlhrskkiklagtlrdklhqvllqpygqeglsk 426
QY 481 SVOLNGOPLVMVDGTLPELKPRLRAGRTLVIPTVMGFVVKVNNALACRYR 534
DB 427 svqlngqplvmvdgtlpeklprlragrtlviptvmgffvkvnnalacryr 480

RESULT 9
ID AAB85217
XX AAB85217 standard; Protein; 480 AA.
AC AAB85217;
XX
DT 07-SEP-2001 (first entry)
XX Heparanase-like protein Hpa2 splice variant #3.
DE Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;
XX cytotatic; neuroprotective; cerebroprotective; immunosuppressive;
KW antipsoriatic; nootropic; antiinflammatory; antiarthritic; antiasthmatic;
KW antidiabetic; antiarteriosclerotic; vulnerary.
XX Homo sapiens.
OS
XX WO200146392-A2.
PN 28-JUN-2001.
XX 21-DEC-2000; 2000WO-GB04963.
XX 22-DEC-1999; 99GB-0030392.
PR 07-APR-2000; 2000GB-0008713.
XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX McKenzie EA, Stamps AC, Terrett JA, Tyson KL;
PI WPI; 2001-418056/44.
DR N-PSDB; AAH22673.
DR Novel homologs of heparanase, present in three splice variants, useful
PT for identifying agents that modulate heparanase, useful in the
PT treatment and/or prophylaxis of abnormal levels of heparanase
XX Claim 1; Fig 3; 97pp; English.
XX The invention provides a homologue to heparanase which is present in
CC three splice variants. The heparanase homologue polypeptide is useful in
CC the treatment of a human or non-human animal or for use in diagnosis.
CC Vectors comprising the heparanase homologue polynucleotides are useful in
CC the transformation or transfection of a prokaryotic or eukaryotic host.
CC The modulators of the polypeptide are useful in the manufacture of a
CC medicament for the treatment and/or prophylaxis of a condition/disease
CC associated with abnormal levels of the heparanase homologue, including
CC cancer, central nervous system (CNS) and neurodegenerative diseases,
CC cardiovascular diseases such as stenosis following angioplasty and
CC atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus,
CC allografts, inflammatory diseases, arthritis, vascular stenosis,
CC tumour growth and progression, asthma, Alzheimer's disease, diabetic
CC retinopathy, wound healing and inflammation. The polypeptide is also
CC useful in diagnosis and research. The present sequence represents the
CC amino acid sequence of the smallest splice variant of the heparanase-
CC like protein Hpa2 of the invention.
XX Sequence 480 AA;
SQ

Query Match 89.0%; Score 2478; DB 22; Length 480;
Best Local Similarity 89.9%; Pred. No. 1.1e-242;
Matches 480; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

OS Homo sapiens.
PN WO200179253-A1.
XX
PD 25-OCT-2001.
XX
PF 11-APR-2001; 2001WO-US11643.
XX
PR 18-APR-2000; 2000US-198123P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Fiscella M, Shi Y, Ebner R, Ruben SM;
XX
DR WPI; 2001-611720/70.
XX
XX
PT New nucleic acids encoding extracellular matrix polypeptides, for
PT diagnosing, treating, preventing or ameliorating human disorders and
PT disease, such as, autoimmune, hyperproliferative or cardiovascular
PT disorders -
XX
PS Disclosure; Page 13-14; 308pp; English.
XX
XX The invention relates to novel isolated polynucleotides (I) encoding
CC extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by
CC (I) are used to prevent, treat or ameliorate a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility to
CC a pathological condition. The antibodies to the polypeptides can also be
CC used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities. The present sequence represents the amino acid sequence
CC of human heparanase-like protein.
XX
XX Sequence 439 AA;

Query Match 81.5%; Score 2270; DB 22; Length 439;
Best Local Similarity 89.0%; Pred. NO. 1.4e-221;
Matches 439; Conservative 0; Mismatches 0; Indels 54; Gaps 1;
QY 42 GDRRLPVDRAAGLEKTKILLDVSSTKNPVRVNVNENFLSLQDPSLIHDGWLDFLSKKRL 101
Db 1 gdrpvpdraaglkextkllldvstknprvtvnenflslqldpslihdgwlfdlsskrl 60
QY 102 VTLAGLSPAFIRFGGKRTDFLQFNLRNPAKSRGPGPDYLYLKNYEDIVRSVDALDKQ 161
Db 61 vtlaglspafirfggkrtdfllqfnlrnpaksrpgpgpdylylknedy----- 108
QY 162 KCKTAQHPDVMLELQREKAAQMHVLLVLLKEQFSNTYSNLIETPNRYTMHGRAVNGSQL 221
Db 109 -----epnnyrtmhgravnsgsl 126
QY 222 GKDYIQLKSLLOPIRYSPASLYGNIGPRKNVIALLDGFMKVGASTYDVTWQHICYD 281
Db 127 gkdyiqklsllqpirysrasylygniprknvialldgfmkvagstvdvtwqhicyid 186
QY 282 GRVVKVMOFLKRLDLSQIRKIQKVNTYTPGKKIWLEGVVTTTSAGGTNNLSDSYAA 341
Db 187 grvkvmdfklrlldtsdqirkikqvntytpgkkiwlegvvtttsaggtnnlstdsyaa 246

QY 342 GFELWLTIGMLANQGDIVVIRHSFDFHGVNHLVDQNFNLPDYLWLSLLYKRLIGPKVLAV 401
Db 247 gflwltlglmlanqgidvvirhsffdhgvnhlvdqnfnpdpdywlsllykrligpkvlav 306
QY 402 HVAGLQKRPGRVIRDKLRIYAHCTNHHNHNHVVRGSIITLFIINLHRSRKKIKLACTLRD 461
Db 307 hvaglqrkprgrvirrdkrlriyahctnnhnhnhyvrgsifcfinlhrsrrkkiklagctlr 366
QY 462 KLVHGYLLQPYGOEGLKSKSVQLNGQPLVMVDDGTLPKLPRLRAGRTLVIPTVTMGFF 521
Db 367 klvhgyllqpyggeglksksvqlngqplvmvddgtlpelkprlragrtlviptvtmgff 426
QY 522 VVKNVNALACRYR 534
Db 427 vvkvnvalacryr 439
RESULT 12
AAM24147
ID AAM24147 standard; Protein; 262 AA.
XX
AC AAM24147;
XX
DT 12-OCT-2001 (first entry)
XX
DE Human EST encoded protein SEQ ID NO: 1672.
XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition.
XX
OS Homo sapiens.
XX WO200154477-A2.
PN
PD 02-AUG-2001.
XX
XX 25-JAN-2001; 2001WO-US02687.
XX
XX 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR WPI; 2001-476164/51.
DR N-PSDB; AAH98806.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
PS Claim 20; Page 1122-1123; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention.
XX
SQ Sequence 262 AA;

Query Match 47.2%; Score 1314; DB 22; Length 262;
Best Local Similarity 98.5%; Pred. NO. 9.1e-125;
Matches 257; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY	1	MRVLCAPPEAMPSSNRRPACIAPGALYIALLLHLSLSSQAGDRPLPVDRAAGLKEKTL	60	PR	14-AUG-2000;	2000US-0225268.
Db	1	mrvlcfafeampssnrrpaciapgalylallllhslssqagdrplpvdraaglkektl	60	PR	14-AUG-2000;	2000US-0225270.
QY	61	ILLDVSTKPNVTNENFLSLQLDPSIIHDGWLDFLSKRLVTLARGLSPAFLRFGGKRT	120	PR	14-AUG-2000;	2000US-0225447.
Db	61	illdvstknprtvenenflslqldpsilhdgwlflsskrlvtlrglspafirfggkrt	120	PR	14-AUG-2000;	2000US-0225757.
QY	121	DFLQFQNLNPAKSRGGPGPDYVLKNYEDDIIVRSVDALDKGCKIAQHPDVMLELQREK	180	PR	14-AUG-2000;	2000US-0225758.
Db	121	dflqfnltrapaksrgppdyvlknyeddiivrsdvaldkgckiaqhpdmlepprek	180	PR	14-AUG-2000;	2000US-0225759.
QY	181	AAQMHVLVLEQPSNTYSNLIITPEPNYRTMHGRAVNSQLGKDYIQLKSILQPIRIYSR	240	PR	14-AUG-2000;	2000US-0226279.
Db	181	aagmhvlvlkeqfsntysnliitpepnnyrtmhgravnsgldgkdyiqklsllqpiirysr	240	PR	14-AUG-2000;	2000US-0226279.
QY	241	ASLYGNIGRPRKNVTIALLDG	261	PR	14-AUG-2000;	2000US-0226681.
Db	241	aslygnivirprknvtialldg	261	PR	14-AUG-2000;	2000US-0226681.
RESULT	13					
AAM99905						
ID	AAM99905	standard; Protein; 214 AA.				
AC	AAM99905;					
XX						
DT	07-JAN-2002	(first entry)				
XX						
DE	Human excretory related polypeptide SEQ ID NO 642.					
XX						
KW	Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;					
KW	immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;					
KW	antiparkinsonian; antisickling; antianaemic; antiarthritis; cancer;					
KW	antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;					
KW	antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;					
KW	antiparasitic; cardiatic; immune disorder; cardiovascular disorder;					
KW	neurological disease; infection; nephrotropic; gene therapy; vaccine;					
KW	excretory system.					
XX						
OS	Homo sapiens.					
XX						
PN	WO200155313-A2.					
XX						
PD	02-AUG-2001.					
XX						
PF	17-JAN-2001; 2001WO-US01323.					
XX						
PR	31-JAN-2000; 2000US-0179065.					
PR	04-FEB-2000; 2000US-0180628.					
PR	24-FEB-2000; 2000US-0184664.					
PR	02-MAR-2000; 2000US-0186350.					
PR	16-MAR-2000; 2000US-0189874.					
PR	17-MAR-2000; 2000US-0190076.					
PR	18-APR-2000; 2000US-0198123.					
PR	19-MAY-2000; 2000US-0205515.					
PR	07-JUN-2000; 2000US-0209467.					
PR	28-JUN-2000; 2000US-0214886.					
PR	30-JUN-2000; 2000US-0215135.					
PR	07-JUL-2000; 2000US-0216647.					
PR	07-JUL-2000; 2000US-0216880.					
PR	11-JUL-2000; 2000US-0217487.					
PR	11-JUL-2000; 2000US-0217496.					
PR	14-JUL-2000; 2000US-0218290.					
PR	26-JUL-2000; 2000US-0220963.					
PR	26-JUL-2000; 2000US-0220964.					
PR	14-AUG-2000; 2000US-0224518.					
PR	14-AUG-2000; 2000US-0224519.					
PR	14-AUG-2000; 2000US-0225213.					
PR	14-AUG-2000; 2000US-0225214.					
PR	14-AUG-2000; 2000US-0225266.					
PR	14-AUG-2000; 2000US-0225267.					

PR	14-AUG-2000;	2000US-0225268.
PR	14-AUG-2000;	2000US-0225270.
PR	14-AUG-2000;	2000US-0225447.
PR	14-AUG-2000;	2000US-0225757.
PR	14-AUG-2000;	2000US-0225758.
PR	14-AUG-2000;	2000US-0225759.
PR	14-AUG-2000;	2000US-0226279.
PR	14-AUG-2000;	2000US-0226681.
PR	22-AUG-2000;	2000US-0226688.
PR	22-AUG-2000;	2000US-0227182.
PR	22-AUG-2000;	2000US-0227009.
PR	23-AUG-2000;	2000US-0228924.
PR	30-AUG-2000;	2000US-0228924.
PR	01-SEP-2000;	2000US-0229287.
PR	01-SEP-2000;	2000US-0229343.
PR	01-SEP-2000;	2000US-0229344.
PR	01-SEP-2000;	2000US-0229345.
PR	05-SEP-2000;	2000US-0229509.
PR	05-SEP-2000;	2000US-0229513.
PR	06-SEP-2000;	2000US-0230437.
PR	06-SEP-2000;	2000US-0230438.
PR	08-SEP-2000;	2000US-0231242.
PR	08-SEP-2000;	2000US-0231243.
PR	08-SEP-2000;	2000US-0231244.
PR	08-SEP-2000;	2000US-0231413.
PR	08-SEP-2000;	2000US-0231414.
PR	08-SEP-2000;	2000US-0232080.
PR	08-SEP-2000;	2000US-0232081.
PR	12-SEP-2000;	2000US-0231968.
PR	14-SEP-2000;	2000US-0232397.
PR	14-SEP-2000;	2000US-0232398.
PR	14-SEP-2000;	2000US-0232399.
PR	14-SEP-2000;	2000US-0232400.
PR	14-SEP-2000;	2000US-0232401.
PR	14-SEP-2000;	2000US-0233063.
PR	14-SEP-2000;	2000US-0233064.
PR	14-SEP-2000;	2000US-0233065.
PR	21-SEP-2000;	2000US-0234223.
PR	21-SEP-2000;	2000US-0234274.
PR	25-SEP-2000;	2000US-0234997.
PR	25-SEP-2000;	2000US-0234998.
PR	26-SEP-2000;	2000US-0235484.
PR	27-SEP-2000;	2000US-0235834.
PR	27-SEP-2000;	2000US-0235836.
PR	29-SEP-2000;	2000US-0236327.
PR	29-SEP-2000;	2000US-0236367.
PR	29-SEP-2000;	2000US-0236368.
PR	29-SEP-2000;	2000US-0236369.
PR	29-SEP-2000;	2000US-0236370.
PR	02-OCT-2000;	2000US-0236802.
PR	02-OCT-2000;	2000US-0237037.
PR	02-OCT-2000;	2000US-0237038.
PR	02-OCT-2000;	2000US-0237039.
PR	02-OCT-2000;	2000US-0237040.
PR	13-OCT-2000;	2000US-0239935.
PR	13-OCT-2000;	2000US-0239937.
PR	20-OCT-2000;	2000US-0240960.
PR	20-OCT-2000;	2000US-0241221.
PR	20-OCT-2000;	2000US-0241785.
PR	20-OCT-2000;	2000US-0241786.
PR	20-OCT-2000;	2000US-0241787.
PR	20-OCT-2000;	2000US-0241808.
PR	20-OCT-2000;	2000US-0241809.
PR	20-OCT-2000;	2000US-0241826.
PR	01-NOV-2000;	2000US-0244617.
PR	01-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465569/50.
XX N-PSDB; AA198878.
XX
XX Isolated nucleic acid molecule encoding excretory system antigen is
XX used in preventing, treating or ameliorating a medical condition -
XX
XX Claim 11; SEQ ID NO 642; 574pp + Sequence Listing; English.
XX
XX The invention relates to novel excretory system related human
XX polynucleotides (AA198567-AA199503) and the encoded proteins
XX (AA199594-AA199913) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy, especially
XX disorders related to the excretory system. The genes are isolated
XX from a range of human tissues disclosed in the specification. The
XX nucleic acids, proteins, antibodies and (ant)agonists are useful in
XX the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemias;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 214 AA;

Query Match 39.9%; Score 1112; DB 22; Length 214;

Best Local Similarity 100.0%; Pred. No. 2.3e-104; Indels 0; Gaps 0;
Matches 210; Conservative 0; Mismatches 0;

Qy 325 VTTSAGTNNLSDSYAAGFLWLTGLMLANOGIDVVIRHSFFDHGYNHLVDQNFNPLPDY 384
Db 5 vttsaggtnnlsdsyaagflwltglmlanogidvvirhsffdhgynhlvdqnfnpdpdy 64
Qy 385 WLSLLYKRLIGPKVLAVHVAGLQKRPGRVIRDKLRIYAHCTNHHNNHNYVSGSTLFLII 444
Db 65 wlslllykrligpkvlavhvaglqrkprgrvirdklriyahctnhhnnhnyvrgstlflfi 124
Qy 445 NLHRSRKIKLAGTLRDLKLVHQLQPYGOEGLKSKVOLNGQPLVMVDDGTLPELKPRP 504
Db 125 nlhrsrrkkiktagtlrdklkvhqllyqpyqeglkksvqlngqplvmvddgtlpelekprp 184
Qy 505 LRAGRTLVIPTVMGFFVVKVNNALACRYR 534
Db 185 lragrtlvipvtmgffvvkvnvalacryr 214

RESULT 14

AA043704
ID AA043704 standard; Protein; 214 AA.

AC AA043704;

XX 24-OCT-2001 (first entry)

XX Human bladder antigen, SEQ ID NO: 98.

XX Human; bladder antigen; cytostatic; immunosuppressive; nootropic;
XX neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;
XX antiinflammatory; antitumor; vulnery; anticonvulsant; antibacterial;
XX antifungal; antiparasitic; cardiant; gene therapy; cancer;
XX immune disorder; cardiovascular disorder; wound healing; infection;
XX neurological disease.

OS Homo sapiens.

XX WO200159064-A2.

XX 16-AUG-2001.

XX 17-JAN-2001; 2001WO-US01342.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 11-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217496.

PR 14-AUG-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235634.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0246174.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-514652/56.
XX N-PSDB; AAI64065.
DR
XX Forty five bladder related polynucleotides, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases -
XX
PS Claim 11; SEQ ID NO 98; 482pp + sequence listing; English.
XX
XX The invention relates to forty five novel bladder related
CC polynucleotides. The polynucleotides and the polypeptides that they
CC encode are useful in the diagnosis, treatment and prevention of:
CC cancer, particularly breast and ovarian cancer, and other cancers
CC of the adrenal gland, bone, bone marrow, breast, gastrointestinal
CC tract, liver, lung, or urogenital system; immune disorders such as
CC Addison's disease, allergies, autoimmune haemolytic anaemia,
CC autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple
CC sclerosis, rheumatoid arthritis and ulcerative colitis; cardiovascular
CC disorders such as myocardial ischaemias; wound healing; neurological
CC diseases such as cerebral anoxia and epilepsy; and infectious diseases
CC such as viral, bacterial, fungal and parasitic infections. Numerous
CC examples of each type of disorder are given in the specification.
CC The polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. The polynucleotides are
CC useful for chromosome identification. They are also useful as probes
CC for diagnosing or treating a disorder related to the female reproductive
CC system, particularly breast and/or ovary cancer. The present
CC sequence is a novel bladder antigen provided in the invention.
CC
CC Note: The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 214 AA;

Query Match 39.9%; Score 1112; DB 22; Length 214;
Best Local Similarity 100.0%; Pred. No. 2.3e-104;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 VTTSGAGTNNLSDSYAAGFLWNTLGLMLANGIDVIRHSFEDHCYHNLVDOENPLPDY 384
DB 5 vttsgaggnnlsdsyaagflwntlgmlangidvirsffdhgynhlvdqgnfplpd 64
QY 385 WLSLLYRLIGPKVLAVHAGLQRPGRVIRDKLRIYAHCTNHNHNYVRGSTITLFI 444
DB 65 wlslllyrligpkvlavhvaglrkprgrvirdklriyahctnhnhnyvrgsiti 124
QY 445 NLHSRKKIKLAGTLRDKLVHGYLLQPGQGLKSKSVQLANGQPLVMYDDGTLPDLKPRP 504
DB 125 nlhsrkkiklagtlrdklvhyllqpygqglksksvqlngqplvmvddgtlpelkprp 184
QY 505 LRAGRTLVPVMTGMFFVKNVNALACRYR 534
DB 185 lragrtlvpvmtgmffvknvnalacryr 214

RESULT 15
AAB08851
ID AAB08851 standard; Protein; 535 AA.
AC AAB08851;
XX
XX
DT 15-JAN-2001 (first entry)
DE A murine heparanase polypeptide.
XX
KW Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
KW heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW wound healing; infection; burn; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease;
KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease.
XX
OS Mus sp.
XX
XX WO2000052178-A1.
XX
XX
XX 08-SEP-2000.
XX
XX 14-FEB-2000; 2000WO-US03542.
XX
XX 01-MAR-1999; 99US-0258892.
XX
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX (FRIE/) FRIEDMAN M M.
XX
XX Pecker I, Vlodaysky I, Feinstein E;
XX
XX WPI: 2000-579289/54.
XX N-PSDB; AAA75081.
XX
XX New polynucleotides encoding a polypeptide having heparanase activity,
XX useful in wound healing and in gene therapy, particularly in treating
XX tumour, inflammation, autoimmunity, neurodegenerative diseases
XX
XX Claim 22; Page 144-145; 152pp; English.
XX
XX The present sequence represents murine protein with heparanase catalytic
XX activity. The heparanase (hpa) polynucleotide is useful in gene therapy,
XX particularly in treating tumour, inflammation or autoimmunity.
XX Particularly, the polynucleotide is useful in modulating the
XX bioavailability of heparin-binding growth factors, cellular responses

CC to heparin-binding growth factors (e.g. bFGF) and cytokines
CC (e.g. interleukin (IL)-8), cell interaction with plasma lipoproteins,
CC cellular susceptibility to certain viral and some bacterial and protozoa
CC infections, or disintegration of neurodegenerative plaques. The
CC polynucleotide is also useful in wound healing (e.g. thermal, chemical
CC or radiation burns), and in the treatment of angiogenesis, restenosis,
CC atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-
CC Strausler Syndrome or Creutzfeldt-Jakob disease), and some viral,
CC bacterial or protozoa infections.
XX
SQ Sequence 535 AA;

Query Match 33.9%; Score 944.5; DB 21; Length 535;
Best Local Similarity 40.0%; Pred. No. 1.1e-86;
Matches 211; Conservative 76; Mismatches 156; Indels 85; Gaps 9;

QY 63 LDVSTKNPVRTVNENFLSLQIDPSIIHD-GWLDFLSRRRLVTLARGLSPAFRPGGKRTD 121
DB 33 ldytkrplrspsflsitidaslatdprflfigsprlralargspaylrfggktld 92
QY 122 FLQFQNLNRPKSRGGPGPDYVLKNYEDDIVRSDVALDKQCKIAQHPDVMLEQREKA 181
DB 93 fliflflf----dpdkeptseersykwqvnhdicrsepy-----saavlrkqvewp 137
QY 182 AQMHLVLLKEQF-----SNTYS-----NLIL----- 202
DB 138 fq-elllliregqkfkstysrsvdmlysfakcsgldlfiglnallrtptdlrnssna 196
QY 203 -----TEPNYRTMHGRAVNSQGLKDYIQLKSLLOPRTIYRSASL 243
DB 197 qlldycsskgylniswelgnepnstfkkahllidqligedfvelhklqr-safqnaki 255
QY 244 YGPNIQRPRKNVIALLDGFMKVAGSTVDAVTWQHICYIDGRVVKVMDFLKTRLLDLSDOI 303
DB 256 yspdigprgktcklrsfrikaggevidstwhhyingriatkeofissdaldtflsv 315
QY 304 RKIQKVVNTYTPGKKIWLEGVVTTTSAGGTNNLSDSYAAGFLWNTLGLMLANGIDVIRH 363
DB 316 qkilkvtkeitpgkvwlgetsayggapllsntfaagfmwldklgisagmgievvmrq 375
QY 364 SFFDHGYNHLVDQNFNPLPDYWLWLLYKRLIGPKVLAVHAGLQRPGRVIRDKLRIY 423
DB 376 vffgagnyhlvdnfeipdywlsllfkklgprvlrsrvkqpd-----rsklrvy 426
QY 424 AHCTNHHNHNHYVRGSTITLFIINLHRSRKKIKLAGTLRDKLVHGYLLQPGQGLKSKSVQ 483
DB 427 lhctnavyhpnyggeditlyvnhvthkikvppplfrkpvdtlylkpspgdgllsksvq 486
QY 484 LNGQPLVMYDDGTLPDLKPRPLRAGRTLVIPTMTGMFFVKNVNALAC 531
DB 487 lngqlkmvdeqtlpalteklpagsalslpafsygffvirknakiac 534

Search completed: July 30, 2002, 08:13:59
Job time: 336 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:13:59 ; Search time 141.47 Seconds
(without alignments)
69.092 Million cell updates/sec

Title: US-09-836-461-2_COPY_42_129

Perfect score: 448

Sequence: 1 GDRRLPVDRAGLKEKTLI.....PAFLRFGKRTDFLQFQNLK 88

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq_032802:**

- 1: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1980.DAT:**
- 2: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1981.DAT:**
- 3: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1982.DAT:**
- 4: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1983.DAT:**
- 5: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1984.DAT:**
- 6: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1985.DAT:**
- 7: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1986.DAT:**
- 8: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1987.DAT:**
- 9: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1988.DAT:**
- 10: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1989.DAT:**
- 11: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1990.DAT:**
- 12: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1991.DAT:**
- 13: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1992.DAT:**
- 14: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1993.DAT:**
- 15: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1994.DAT:**
- 16: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1995.DAT:**
- 17: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1996.DAT:**
- 18: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1997.DAT:**
- 19: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1998.DAT:**
- 20: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1999.DAT:**
- 21: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA2000.DAT:**
- 22: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA2001.DAT:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	448	100.0	262	22	AA1980.DAT
2	448	100.0	439	22	AAU07423
3	448	100.0	480	22	AAU07418
4	448	100.0	480	22	AA1985.DAT
5	448	100.0	480	22	AA197634
6	448	100.0	534	22	AA1985.DAT
7	448	100.0	538	22	AA197633
8	448	100.0	592	22	AAU07424
9	448	100.0	592	22	AA1985.DAT
10	448	100.0	592	22	AA1985.DAT
11	448	100.0	592	22	AA197632

12	442	98.7	492	22	AA1985.DAT
13	166.5	37.2	535	21	AA1985.DAT
14	150.5	33.6	532	20	AA197083
15	150.5	33.6	543	20	AA197082
16	150.5	33.6	543	20	AA197082
17	150.5	33.6	543	21	AA1985.DAT
18	150.5	33.6	543	21	AA1985.DAT
19	150.5	33.6	543	21	AA1985.DAT
20	150.5	33.6	543	22	AA1985.DAT
21	150.5	33.6	543	22	AA1985.DAT
22	150.5	33.6	543	22	AA1985.DAT
23	150.5	33.6	543	22	AA1985.DAT
24	150.5	33.6	588	20	AA1985.DAT
25	150.5	33.6	592	20	AA1985.DAT
26	149.5	33.4	592	20	AA1985.DAT
27	93	20.8	32	20	AA1985.DAT
28	84.5	18.9	488	22	AA1985.DAT
29	82.5	18.4	488	22	AA1985.DAT
30	80.5	18.0	488	22	AA1985.DAT
31	76.5	17.1	488	22	AA1985.DAT
32	72.5	16.2	280	21	AA1985.DAT
33	72.5	16.2	280	21	AA1985.DAT
34	72.5	16.2	302	21	AA1985.DAT
35	72.5	16.2	302	21	AA1985.DAT
36	72.5	16.2	304	21	AA1985.DAT
37	72.5	16.2	304	21	AA1985.DAT
38	68.5	15.3	211	21	AA1985.DAT
39	64	14.3	349	21	AA1985.DAT
40	64	14.3	369	21	AA1985.DAT
41	64	14.3	369	21	AA1985.DAT
42	64	14.3	405	21	AA1985.DAT
43	64	14.3	2273	17	AA1985.DAT
44	63.5	14.2	935	22	AA1985.DAT
45	63.5	14.2	1306	19	AA1985.DAT

ALIGNMENTS

RESULT	1
AA1985.DAT	
ID	AA1985.DAT standard; Protein; 262 AA.
AC	AA1985.DAT
DT	12-OCT-2001 (first entry)
DE	Human EST encoded protein SEQ ID NO: 1672.
KW	Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW	Tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW	diagnostics; forensic test; gene mapping; genetic disorder;
KW	biodiversity; gene therapy; nutrition.
OS	Homo sapiens.
PN	WO200154477-A2.
PD	02-AUG-2001.
PF	25-JAN-2001; 2001WO-US02687.
PR	25-JAN-2000; 2000US-0491404.
PR	17-JUL-2000; 2000US-0617746.
PR	03-AUG-2000; 2000US-0631451.
PR	15-SEP-2000; 2000US-0663870.
XX	(HYSE-) HYSEQ INC.
XX	Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
XX	Cao Y, Drmanac RA, Zhang J, Werhman T;
XX	WPI; 2001-476164/51.

DR N-PSDB; AAH98806.
XX Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
XX
XX Claim 20; Page 1122-1123; 1275pp; English.
XX
XX The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention.
XX
SQ Sequence 262 AA;

Query Match 100.0%; Score 448; DB 22; Length 262;
Best Local Similarity 100.0%; Pred. No. 2.3e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GDRPLPVDRAAGLKEKTLILLDVSTKNPVTNENFLSLQLDPSIIHDGWLDFLSSKRL 60
Db 42 GDRPLPVDRAAGLKEKTLILLDVSTKNPVTNENFLSLQLDPSIIHDGWLDFLSSKRL 101

Qy 61 VTLAGLSPAFRLPGGKRTDFLQFNLR 88
Db 102 VTLAGLSPAFRLPGGKRTDFLQFNLR 129

RESULT 2
AAU07423
ID AAU07423 standard; Protein: 439 AA.
XX AC AAU07423;
XX
XX 18-DEC-2001 (first entry)
XX Human heparanase-like protein.
XX
XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW antiproliferative; cardiant; vasotropic; cerebroprotective; neurotropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;
KW nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;
KW wound healing; food additive; heparanase.
XX
XX Homo sapiens.
XX
XX WO200179253-A1.
XX
XX 25-OCT-2001.
XX
XX 11-APR-2001; 2001WO-US11643.
XX
XX 18-APR-2000; 2000US-198123P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Fiscella M, Shi Y, Ebner R, Ruben SM;
XX
XX WPI; 2001-611720/70.
XX
XX New nucleic acids encoding extracellular matrix polypeptides, for
PT diagnosing, treating, preventing or ameliorating human disorders and
PT disease, such as, autoimmune, hyperproliferative or cardiovascular
PT disorders -
XX
XX Disclosure; Page 13-14; 308pp; English.

XX The invention relates to novel isolated polynucleotides (1) encoding
CC extracellular matrix (ECM) polypeptides. (1) and a polypeptide encoded by
CC (1) are used to prevent, treat or ameliorate a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility to
CC a pathological condition. The antibodies to the polypeptides can also be
CC used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities. The present sequence represents the amino acid sequence
CC of human heparanase-like protein.
XX
SQ Sequence 439 AA;

Query Match 100.0%; Score 448; DB 22; Length 439;
Best Local Similarity 100.0%; Pred. No. 4.4e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GDRPLPVDRAAGLKEKTLILLDVSTKNPVTNENFLSLQLDPSIIHDGWLDFLSSKRL 60
Db 1 GDRPLPVDRAAGLKEKTLILLDVSTKNPVTNENFLSLQLDPSIIHDGWLDFLSSKRL 60

Qy 61 VTLAGLSPAFRLPGGKRTDFLQFNLR 88
Db 61 VTLAGLSPAFRLPGGKRTDFLQFNLR 88

RESULT 3
AAU07418
ID AAU07418 standard; Protein: 480 AA.
XX AC AAU07418;
XX
XX 18-DEC-2001 (first entry)
XX
XX Novel human extracellular matrix (ECM) protein #1.
XX
XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW antiproliferative; cardiant; vasotropic; cerebroprotective; neurotropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;
KW nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;
KW wound healing; food additive.
XX
XX Homo sapiens.
XX
XX WO200179253-A1.
XX
XX 25-OCT-2001.
XX
XX 11-APR-2001; 2001WO-US11643.
XX
XX 18-APR-2000; 2000US-198123P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Fiscella M, Shi Y, Ebner R, Ruben SM;
XX
XX

DR WPI: 2001-611720/70.
DR N-PSDB; AAS13843.
XX
PT New nucleic acids encoding extracellular matrix polypeptides, for
PT diagnosing, treating, preventing or ameliorating human disorders and
PT disease, such as, autoimmune, hyperproliferative or cardiovascular
PT disorders -
XX
XX Claim 1; Page 292-293; 308pp; English.
XX
XX The invention relates to novel isolated polynucleotides (I) encoding
CC extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by
CC (I) are used to prevent, treat or ameliorate a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility to
CC a pathological condition. The antibodies to the polypeptides can also be
CC used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities. The present sequence represents the amino acid sequence
CC of novel human extracellular matrix (ECM) protein #1.
XX
SQ Sequence 480 AA;

Query Match 100.0%; Score 448; DB 22; Length 480;
Best Local Similarity 100.0%; Pred. No. 4.9e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHGWDLFLSSKRL 60
Db 42 gdrpripvdraaglkeltllldvstknprtvtvnenflslqldpsiihdgwdflsskrl 101

QY 61 VTLAGLSPAFLRFGGKRTDFLQFNLR 88
Db 102 vtlarglspafirfggkrtdfiqfnlr 129

RESULT 4
AAB85217
ID AAB85217 standard; Protein; 480 AA.
XX
AC AAB85217;
XX
DT 07-SEP-2001 (first entry)
XX
DE Heparanase-like protein Hpa2 splice variant #3.
XX
KW Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;
KW cytosolic; neuroprotective; cerebroprotective; immunosuppressive;
KW antipsoriatic; nootropic; antiinflammatory; antiarthritic; antiasthmatic;
KW antidiabetic; antiarteriosclerotic; vulneryary.
XX
OS Homo sapiens.
XX
PN WO200146392-A2.
XX
PD 28-JUN-2001.
XX
XX 21-DEC-2000; 2000WO-GB04963.
XX
XX 22-DEC-1999; 99GB-0030392.
PR

PR 07-APR-2000; 2000GB-0008713.
XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
PA
XX McKenzie EA, Stamps AC, Terrett JA, Tyson KL;
PI
XX WPI: 2001-418056/44.
DR N-PSDB; AAH22673.
XX
XX Novel homologs of heparanase, present in three splice variants, useful
PT for identifying agents that modulate heparanase, useful in the
PT treatment and/or prophylaxis of abnormal levels of heparanase -
XX
XX Claim 1; Fig 3; 97pp; English.
XX
XX The invention provides a homologue to heparanase which is present in
CC three splice variants. The heparanase homologue polypeptide is useful in
CC the treatment of a human or non-human animal or for use in diagnosis.
CC Vectors comprising the heparanase homologue polynucleotides are useful in
CC the transformation or transfection of a prokaryotic or eukaryotic host.
CC The modulators of the polypeptide are useful in the manufacture of a
CC medicament for the treatment and/or prophylaxis of a condition/disease
CC associated with abnormal levels of the heparanase homologue, including
CC cancer, central nervous system (CNS) and neurodegenerative diseases,
CC cardiovascular diseases such as restenosis following angioplasty and
CC atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus,
CC allografts, inflammatory diseases, arthritis, vascular restenosis,
CC tumour growth and progression, asthma, Alzheimer's disease, diabetic
CC retinopathy, wound healing and inflammation. The polypeptide is also
CC useful in diagnosis and research. The present sequence represents the
CC amino acid sequence of the smallest splice variant of the heparanase-
CC like protein Hpa2 of the invention.
XX
SQ Sequence 480 AA;

Query Match 100.0%; Score 448; DB 22; Length 480;
Best Local Similarity 100.0%; Pred. No. 4.9e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHGWDLFLSSKRL 60
Db 42 gdrripvdraaglkeltllldvstknprtvtvnenflslqldpsiihdgwdflsskrl 101

QY 61 VTLAGLSPAFLRFGGKRTDFLQFNLR 88
Db 102 vtlarglspafirfggkrtdfiqfnlr 129

RESULT 5
AAY97634
ID AAY97634 standard; Protein; 480 AA.
XX
AC AAY97634;
XX
DT 20-APR-2001 (first entry)
XX
DE Human heparanase, hnhp1 pn5 form, protein sequence.
XX
KW Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;
KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
KW gene therapy; human.
XX
OS Homo sapiens.
XX
PN WO200100643-A2.
XX
PD 04-JAN-2001.
XX
XX 19-JUN-2000; 2000WO-IL00358.
XX
XX 25-JUN-1999; 99US-0140801.
PR

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX Pecker I, Michal I, Itzhaki H;
XX WPI; 2001-137930/14.
XX N-PSDB; AAA91099.
XX New polynucleotides and polypeptides that are distantly homologous to
PT heparanase, useful in wound healing, as well as in gene therapy
PT protocols for angiogenesis, restenosis, atherosclerosis, or
PT inflammation -
XX Claim 10; Page 63; 67pp; English.
XX This sequence represents a heparanase of the invention.
CC The heparanase DNA and protein sequences are useful in wound healing,
CC angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary
CC diseases, neurodegenerative diseases (such as Scrape, Alzheimer's
CC disease, and Creutzfeldt-Jakob disease) or viral infections. The
CC heparanase coding sequence is particularly useful in gene therapy.
XX
SQ Sequence 480 AA;
Query Match 100.0%; Score 448; DB 22; Length 480;
Best Local Similarity 100.0%; Pred. No. 4.9e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHGWLDLSSKRL 60
Db 42 gdrripvdraaglkeltllldvstknprtvtvnenflslqlqpsiihdgwldflsskrl 101
QY 61 VTLAGLSPAFLEFGGKRTDFLOFNLR 88
Db 102 vtlarglspafirfggkrtdfiqfnlr 129
RESULT 6
ID AAB85216 standard; Protein; 534 AA.
XX AAB85216;
XX 07-SEP-2001 (first entry)
XX Heparanase-like protein Hpa2 splice variant #2.
XX Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;
KW cytosolic; neuroprotective; cerebroprotective; immunosuppressive;
KW antipsoriatic; nootropic; antiinflammatory; antiarthritic; antiasthmatic;
KW antidiabetic; antiarteriosclerotic; vulnery.
XX Homo sapiens.
OS
XX WO200146392-A2.
PN
XX 28-JUN-2001.
PD
XX 21-DEC-2000; 2000WO-GB04963.
PF
XX 22-DEC-1999; 99GB-0030392.
PR
XX 07-APR-2000; 2000GB-0008713.
PR
XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
PA
XX McKenzie EA, Stamps AC, Terrett JA, Tyson KL;
PI
XX WPI; 2001-418056/44.
DR
XX N-PSDB; AAH22672.
DR
XX Novel homologs of heparanase, present in three splice variants, useful
PT for identifying agents that modulate heparanase, useful in the

PT treatment and/or prophylaxis of abnormal levels of heparanase -
XX Claim 1; Fig 2; 97pp; English.
XX The invention provides a homologue to heparanase which is present in
CC three splice variants. The heparanase homologue polypeptide is useful in
CC the treatment of a human or non-human animal or for use in diagnosis.
CC Vectors comprising the heparanase homologue polynucleotides are useful in
CC the transfection or transfection of a prokaryotic or eukaryotic host.
CC The modulators of the polypeptide are useful in the manufacture of a
CC medicament for the treatment and/or prophylaxis of a condition/disease
CC associated with abnormal levels of the heparanase homologue, including
CC cancer, central nervous system (CNS) and neurodegenerative diseases,
CC cardiovascular diseases such as restenosis following angioplasty and
CC atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus,
CC allografts, inflammatory diseases, arthritis, vascular restenosis,
CC tumour growth and progression, asthma, Alzheimer's disease, diabetic
CC retinopathy, wound healing and inflammation. The polypeptide is also
CC useful in diagnosis and research. The present sequence represents the
CC amino acid sequence of the mid-sized splice variant of the heparanase-
XX like protein Hpa2 of the invention.
XX
SQ Sequence 534 AA;
Query Match 100.0%; Score 448; DB 22; Length 534;
Best Local Similarity 100.0%; Pred. No. 5.6e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHGWLDLSSKRL 60
Db 42 gdrripvdraaglkeltllldvstknprtvtvnenflslqlqpsiihdgwldflsskrl 101
QY 61 VTLAGLSPAFLEFGGKRTDFLOFNLR 88
Db 102 vtlarglspafirfggkrtdfiqfnlr 129
RESULT 7
ID AAY97633 standard; Protein; 538 AA.
XX AAY97633;
XX 20-APR-2001 (first entry)
XX Human heparanase, hnhp1 pn9 form, protein sequence.
XX Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;
KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
KW gene therapy; human.
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH Misc-difference 305
FT /note= "encoded by GAC"
XX
XX WO200100643-A2.
PN
XX 04-JAN-2001.
PD
XX 19-JUN-2000; 2000WO-IL00358.
PF
XX 25-JUN-1999; 99US-0140801.
PR
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA
XX Pecker I, Michal I, Itzhaki H;
PI
XX WPI; 2001-137930/14.
DR
XX N-PSDB; AAA91098.

[illegible]

XX WO200100643-A2.
PN 04-JAN-2001.
XX 19-JUN-2000; 2000WO-IL00358.
XX 25-JUN-1999; 99US-0140801.
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA Pecker I, Michal I, Itzhaki H;
PI WPI; 2001-137930/14.
XX N-PSDB; AAA91097.
DR New polynucleotides and polypeptides that are distantly homologous to
XX heparanase, useful in wound healing, as well as in gene therapy
PT protocols for angiogenesis, restenosis, atherosclerosis, or
PT inflammation -
XX Claim 10; Fig 1; 67pp; English.
PS This sequence represents a heparanase of the invention.
XX The heparanase DNA and protein sequences are useful in wound healing,
CC angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary
CC diseases, neurodegenerative diseases (such as Scurvy, Alzheimer's
CC disease, and Creutzfeldt-Jakob disease) or viral infections. The
CC heparanase coding sequence is particularly useful in gene therapy.
XX Sequence 592 AA;
SQ

Query Match 100.0%; Score 448; DB 22; Length 592;
Best Local Similarity 100.0%; Pred. No. 6.4e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GDRPLPVDRAAGLKEKTLILLDVSTKNPVTNENFLSLQLDPSIIHDGWLDFLSKRL 60
|||||
Db 42 gdrplpvdraaglkectllldvstknprtvtnefnflslqlpsihdgdwldflsskrl 101
|||||
Qy 61 VTLARGLSPAFLRFGGKRTDFLOFNLR 88
|||||
Db 102 vtlarglspafllrfggkrtldflqgnlr 129
|||||

RESULT 12
AAB84664
ID AAB84664 standard; Protein; 492 AA.
XX AAB84664;
XX 05-SEP-2001 (first entry)
XX Amino acid sequence of human heparanase-like polypeptide.
DE Human; heparanase-like polypeptide; gene therapy; cancer; angiogenesis;
XX trauma; autoimmune disease; skin disease; cardiovascular disease;
KW nervous system disease; inflammation; arthritis; genitalia;
KW male fertility; erectile dysfunction.
XX Homo sapiens.
OS Key Location/Qualifiers
FH Misc-difference 407
FT /note= "unspecified residue encoded by KCA"
XX WO200148161-A2.
PN 05-JUL-2001.
XX 18-DEC-2000; 2000WO-EP12909.
XX

PR 23-DEC-1999; 99EP-0125831.
XX (SCHD) SCHERING AG.
PA Siemeister G, Weiss B;
XX WPI; 2001-418259/44.
DR N-PSDB; AAH28347.
XX Human Heparanase-like polynucleotide encoding polypeptides useful for
PT modulating expression of the polypeptide and for treating cancer, -
PT cancer metastasis, aberrant angiogenesis by gene therapy technique -
XX Claim 9; Page 30; 30pp; English.
PS The present sequence represents a human heparanase-like polypeptide.
XX Heparanase-like polynucleotides are useful as a source of probes,
CC primers and antisense molecules, and in gene therapy. Heparanase-like
CC polynucleotides and polypeptides are useful for treating several
CC disorders e.g., cancer, cancer metastasis. The oligonucleotides are
CC also useful as diagnostic markers for the diagnosis of disorder such
CC as cancer, cancer metastasis and aberrant angiogenesis. They may also
CC act as diagnostic markers for diagnosis of disorder such as cancer.
CC Cancer metastasis and aberrant angiogenesis. The heparanase polypeptides
CC and polynucleotides are also useful for treating trauma, autoimmune
CC diseases, skin diseases, cardiovascular diseases, nervous system
CC diseases, and inflammation including arthritis. Since the polynucleotide
CC is preferentially expressed in male genitalia, modulation of its
CC expression and/or activity may be used for medical intervention in male
CC genitalia function that is male fertility control, erectile dysfunction.
XX Sequence 492 AA;
SQ

Query Match 98.7%; Score 442; DB 22; Length 492;
Best Local Similarity 100.0%; Pred. No. 2.9e-47;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 DRRLPVDRAAGLKEKTLILLDVSTKNPVTNENFLSLQLDPSIIHDGWLDFLSKRLV 61
|||||
Db 1 drrlpvdraaglkectllldvstknprtvtnefnflslqlpsihdgdwldflsskrlv 60
|||||
Qy 62 TLARGLSPAFLRFGGKRTDFLOFNLR 88
|||||
Db 61 tlarglspafllrfggkrtldflqgnlr 87
|||||

RESULT 13
AAB08851
ID AAB08851 standard; Protein; 535 AA.
XX AAB08851;
XX 15-JAN-2001 (first entry)
XX A murine heparanase polypeptide.
DE Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
XX heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW wound healing; infection; burn; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease;
KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease.
XX Mus sp.
OS WO200052178-A1.
XX 08-SEP-2000.
XX 14-FEB-2000; 2000WO-US03542.
XX 01-MAR-1999; 99US-0258892.
XX


```

XX Polynucleotides encoding mammalian endoglucuronidases, especially
PT heparanases, useful to promote wound healing
XX
XX Claim 6; Page 69-73; 112pp; English.
XX
CC The invention relates to nucleic acid sequences that encode heparanase
CC enzymes having endoglucuronidase activity. Recombinant heparanases are
CC capable of removing the HS side chain from heparan sulfate proteoglycan
CC (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to
CC inhibit heparanase, this is useful for treatment of a physiological or
CC medical condition associated with elevated heparanase activity, such as
CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,
CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and
CC rat heparanases can be used to enhance wound healing, especially
CC associated with tissue development and repair. The conditions mentioned
CC above can be diagnosed using specific antibodies, and also using primers
CC and probes specific for the heparanase polynucleotides. Other uses of the
CC heparanases include sequencing sulfated molecules such as HSPG. The
CC present sequence represents a human heparanase.
XX
XX Sequence 543 AA;
SQ
Query Match 33.6%; Score 150.5; DB 20; Length 543;
Best Local Similarity 43.5%; Pred. No. 2.5e-10;
Matches 37; Conservative 17; Mismatches 28; Indels 3; Gaps 2
QY 1 GDRRLPVDRAAGLKEKTLILLIDSTKNPRTVNENFLSLQIDPSIIHD-GWLDFLSKKR 59
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
22 gplspgalpraa--qagdvdlldfftqehlhpvsflsvtldanlatdprflllgspk 79
QY 60 LVTLARGLSPAFLRFGRKRTDFLOF 84
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
80 lrtlarglspaylrfggktclflf 104
Search completed: July 30, 2002, 08:14:00
Job time: 337 sec

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:16:22 ; Search time 71.68 Seconds
(without alignments)
117.967 Million cell updates/sec

Title: US-09-836-461-2_COPY_42_129
Perfect score: 448
Sequence: 1 GDRRLPVDRAAGLKEKTLI.....PAFLRFGGKRTDFLQFQNL 88
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	448	100.0	480	JC7506	heparanase protein
2	73.5	16.4	253	AF3625	cell division inhi
3	73	16.3	481	S56299	hypothetical prote
4	72.5	16.2	334	A86397	protein T7N9.6 [im
5	68.5	15.3	694	F90609	exodeoxyribonuclea
6	67.5	15.1	293	T09171	ribosomal protein
7	67.5	15.1	293	T09170	ribosomal protein
8	65.5	14.6	1391	T20642	hypothetical prote
9	65.5	14.6	1397	E87998	protein F09C3.1 li
10	65	14.5	2288	T30568	acetyl-CoA carboxy
11	63.5	14.2	492	S71455	catalase (EC 1.11.
12	63.5	14.2	497	T29814	hypothetical prote
13	63	14.1	1178	AC3394	DNA polymerase III
14	62.5	14.0	326	A97316	hypothetical prote
15	62.5	14.0	618	S33044	hypothetical prote
16	62.5	14.0	732	S05238	peptidyl-dipectida
17	62.5	14.0	1306	A31759	stage II sporulati
18	62	13.8	830	F83659	hypothetical prote
19	61.5	13.7	126	E71233	hypothetical prote
20	61.5	13.7	299	S25782	hypothetical prote
21	61.5	13.7	299	D23696	lacX protein - Lac
22	61.5	13.7	302	E70831	hypothetical prote
23	61.5	13.7	1299	AH2090	two-component hybr
24	61	13.6	126	B75200	hypothetical prote
25	61	13.6	275	S52278	hypothetical prote
26	61	13.6	352	D97947	UDP-N-acetylglucos
27	61	13.6	423	T03269	uroporphyrin-III C
28	61	13.6	426	B69455	conserved hypothet
29	61	13.6	433	C82899	thymidine phosphor

RESULT 1
JC7506
heparanase protein 2a - human
C:Species: Homo sapiens (man)
C:Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 01-Dec-2000
C:Accession: JC7506
R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hircok, M.;
Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000
A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase fami
A:Reference number: JC7506
A:Accession: JC7506
A:Molecule type: mRNA
A:Residues: 1-480 <MCK>
A:Cross-references: GB:AF282885
C:Comment: This protein, a intracellular membrane-bound enzyme, has biological and th
therapies.
C:Genetics:
A:Gene: hpa2a
A:Map position: 10q23-10q24
C:Keywords: heparin binding; membrane bound

Query Match 100.0%; Score 448; DB 2; Length 480;
Best Local Similarity 100.0%; Pred. No. 2.8e-42;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GDRRLPVDRAAGLKEKTLIILDVSTKNPVRTVNENFLSLQLDPSIIHDGWLDFLSSRRL 60
|||||
Db 42 GDRRLPVDRAAGLKEKTLIILDVSTKNPVRTVNENFLSLQLDPSIIHDGWLDFLSSRRL 101
|||||

Qy 61 VTLARGLSPAFLRFGGKRTDFLQFQNL 88
|||||
Db 102 VTLARGLSPAFLRFGGKRTDFLQFQNL 129
|||||

RESULT 2
AF3625
cell division inhibitor minC [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C:Accession: AF3625
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanov
; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit
A:Reference number: AD3252; PMID:11756688
A:Accession: AF3625
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-253 <KUR>
A:Cross-references: GB:AF008918; PIDN:AAU54169.1; PID:gl7985135; GSPDB:GN00191
A:Experimental source: strain 16M

ALIGNMENTS

```

C:Genetics:
A:Gene: BMEII0927
A:Map position: II
C:Superfamily: cell division inhibitor minC

Query Match 16.4%; Score 73.5; DB 2; Length 253;
Best Local Similarity 37.3%; Pred. No. 1.2;
Matches 22; Conservative 7; Mismatches 21; Indels 9; Gaps 2;

Qy 29 PVRTVNFSLQLDPSIIHDGWLDFLSKRLVTLARGLSAPFLREGGKRTDFLQFQNL 87
      I:I :||: : : ||||: || ||| | | | | : | : ||
Db 15 FIRLKGSRFLAMVLSPELPLDGWLE-----RLDDLARRSSGGFFL----GRPVVLDMENT 64

```

N: Alternate names: hypothetical protein F006
C: Species: Saccharomyces cerevisiae
C: Date: 02-Sep-1995 #sequence_revision 19-Oct-1995 #text_change 29-Oct-1999

C:Accession: S56299; S62355; S63791
R: Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S. I.; Sasano
submitted to the EMBL Data Library, May 1995
sequence of chromosome VI from *Saccaromyces*

A; Description: Analysis of the nucleotide sequence of Cytomegalovirus
A; Reference number: S56186
A; Accession: S56299
A; Molecule type: DNA
A; Residues: 1-481 <MUR>

A, Cross-references: EMBL:D50617; NID:g836685; PIDN:BAA09283.1; PID:Q1009924; PFD:g836794
R:Murakami, Y.
submitted to the EMBL Data Library, December 1994
A, Reference number: S62230

A.Accession: S62255
A.Molecule type: DNA
A.Residues: 1-481 <NUM>
A.Cross-references: EMBL:D45397; NID:g871938; PIDN:BAA08010.1; PID:d1008600; PID:g871943
R.Eki, T., Naitou, M.; Hagiwara, H.; Ozawa, M.; Sasanuma, S.I.; Sasanuma, M.; Tsuchiya, Y.
Yeast 12, 149-167, 1996
A.Title: Analysis of a 36.2 kb DNA sequence including the right telomere of chromosome V
A.Reference number: S63787; MUID:96287652

A:Accession: S63791
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-481 <EKI>
A:Cross-references: EMBL:D44597; NID:g871938; PID:d1008600; PID:g871943
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1995
C:Genetics:
A:Map position: 6R
A>Note: YPR044c

Query Match	16.3%	Score 73;	DB 2;	Length 481;
Best Local Similarity	28.6%	Pred. No. 3;		
Matches	28;	Conservative	11;	Mismatches 31;
				Indels 28;
				Gaps 4;

	KTLILLDVSTKNPRTV-----NENFLSL-----QLDPSIIHDG-WL--	52
QY	: :	
Dh	KTVIPAKVFGEFSGTIRTPVDMDSEKLTLSIQVKHCDKAFKLSNPNCRTELIIDHGAYWSD	393

QY	53	-----DFLSSKKLVTLARGLSPAFLRFGGKRTDFLQFN	86
		- : : - : : - : : - : : - : :	
Db	394	PFNAQFIAAKKATKLVGVDPDFTREGGSIPITLTFQD	431

RESULT 4
A86397
protein T7N9_6 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: A86397
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Devarkar, R.

A:Cross-references: EMBL:Y15996; NID:g3021302; PIDN:CAA75926.1; PID:g3021303

C:Genetics:

A:Gene: accA

A:Introns: 54/3; 111/3

C:Superfamily: human acetyl-CoA carboxylase; biotin carboxylase homology; lipoyl/biotin-
C:Keywords: ligase

Query Match 14.5%; Score 65; DB 2; Length 2288;

Best Local Similarity 32.1%; Pred. No. 1.6e+02;

Matches 18; Conservative 9; Mismatches 25; Indels 4; Gaps 1;

QY 9 DRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSI-----IHGWLDFLSSKRL 60

Db 507 NRSARKHVMIALKELSGRFRITIEYLKLETPAFENKITTGWLQLISNKL 562

RESULT 11

S71455

Catalase (EC 1.11.1.6) 2 - maize

C:Species: Zea mays (maize)

C:Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 04-Mar-2000

C:Accession: S71455; S18819; A36062

R:Guan, L.; Polidoros, A.N.; Scandalios, J.G.

Plant Mol. Biol. 30, 913-924, 1996

A:Title: Isolation, characterization and expression of the maize Cat2 catalase gene.

A:Reference number: S71455; MUID:96270370

A:Accession: S71455

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-492 <GUA>

A:Cross-references: EMBL:Z54358

A:Note: due to an additional T the nucleic acid sequence submitted to the EMBL Data Lib

A:Note: the authors translated the codon AAC for residue 28 as Met, GCC for residue 241

R:Guan, L.; Ruzsa, S.; Skadsen, R.W.; Scandalios, J.G.

Plant Physiol. 96, 1379-1381, 1991

A:Title: Comparison of the cat2 complementary DNA sequences of a normal catalase activit

A:Reference number: S18819

A:Accession: S18819

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-240,'RC',244-296,'T',298-492 <GUW>

A:Cross-references: EMBL:X54819; NID:g22233; PIDN:CAA38588.1; PID:g22234

R:Bethards, L.A.; Skadsen, R.W.; Scandalios, J.G.

Proc. Natl. Acad. Sci. U.S.A. 87, 6927, 1990

A:Reference number: A36062; MUID:90370897

A:Contents: erratum

A:Accession: A36062

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-123,'N',125-240,'RC',244-296,'R',298-492 <BET>

A:Cross-references: GB:J02976

A:Note: the authors translated the codon AAC for residue 124 as Lys

C:Genetics:

A:Gene: cat2

A:Introns: 5/3; 38/1; 419/3; 442/2; 473/3

C:Superfamily: catalase

C:Keywords: Chromoprotein; heme; iron; metalloprotein; oxidoreductase

F:64,103,137/Active site: His, Ser, Asn #status predicted

F:348/Binding site: heme iron (Tyr) (axial ligand) #status predicted

Query Match 14.2%; Score 63.5; DB 2; Length 492;

Best Local Similarity 32.8%; Pred. No. 36;

Matches 20; Conservative 10; Mismatches 22; Indels 9; Gaps 3;

QY 3 RRLPLVDRAAGLUKTKTLLLDVSTKNPVRTVNENFLSLQLDPS-----IHDGWLDFLSSKR 59

Db 405 RYPIPTAHIAGRREKTVI----SKENNFQKQERYA--MDPARQERTITRWVDAISDP 458

QY 60 L 60

Db 459 L 459

RESULT 12

T29814

hypothetical protein C46A5.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T29814

R:Johnson, D.; Stellyes, L.

submitted to the EMBL Data Library, June 1996

A:Description: The sequence of C. elegans cosmid C46A5.

A:Reference number: Z20690

A:Accession: T29814

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-497 <OH>

A:Cross-references: EMBL:U61948; PIDN:AAB03142.1; GSPDB:GN00022; CESP:C46A5.1

A:Experimental source: strain Bristol N2; clone C46A5

C:Genetics:

A:Gene: CESP:C46A5.1

A:Map position: 4

A:Introns: 85/3; 108/1; 189/1; 221/3; 329/2; 367/2; 444/3; 473/3

Query Match 14.2%; Score 63.5; DB 2; Length 497;

Best Local Similarity 26.0%; Pred. No. 37;

Matches 19; Conservative 8; Mismatches 17; Indels 29; Gaps 2;

QY 43 DPSIIHDGWLDFLSSKRLVTLARGLSA-----FLRFGG-- 76

Db 264 DSSYIHANWVDLNSKKAILTQLPLSHTASDFWQMIIDQIKCVLLIMTDGEFNKFGNS 323

QY 77 ---KRTDFLOFQN 86

Db 324 VFPOQDFLKFED 336

RESULT 13

AC3394

DNA polymerase III, alpha chain (EC 2.7.7.7) [Imported] - Brucella melitensis (strain

C:Species: Brucella melitensis

C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002

C:Accession: AC3394

R:DeIvecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanov

; Mazur, M.; Goitsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit

A:Reference number: AD3252; PMID:11756688

A:Accession: AC3394

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1178 <KUR>

A:Cross-references: GB:AE008917; PIDN:AAL52318.1; PID:gl7983111; GSPDB:GN00190

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BME1137

A:Map position: 1

C:Superfamily: DNA-directed DNA polymerase III alpha chain

C:Keywords: nucleotidyltransferase

Query Match 14.1%; Score 63; DB 2; Length 1178;

Best Local Similarity 31.2%; Pred. No. 1.2e+02;

Matches 24; Conservative 6; Mismatches 43; Indels 4; Gaps 2;

QY 8 VDRAAGLKEKTKTLL--DVSTKNPVRTVNENFLSLQL-DPSIIHDGWLDFLSSKRLVTL 63

Db 112 VNRKALDLAPLVLAATEAGYANIVRLVRAFLDTPASDPIHIEAGWLPALSDVIALT 171

QY 64 ARGLSPAFLREGKRTD 80

Db 172 GGPLGPIGRSFTADRAD 188

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: July 30, 2002, 08:31:25 ; Search time 36.14 Seconds
(without alignments)
94.281 Million cell updates/sec
Title: US-09-836-461-2_COPY_42_129
Perfect score: 448
Sequence: 1 GDRPLPVDRAAGLKEKTLI.....PAFLRFGKRTDQLQFNLR 88
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : SWISSProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	73	16.3	481	1 YFL4_YEAST	P43616 saccharomyc
2	67.5	15.1	293	1 RK4_SPIOL	O49937 spinacia ol
3	63.5	14.2	491	1 CAT2_MAIZE	P12365 zea mays (m
4	63	14.1	219	1 C270_MOUSE	P58468 mus musculu
5	63	14.1	987	1 SOXA_RHIME	O87386 rhizobium m
6	62.5	14.0	575	1 UL87_EBV	P25215 Epstein-bar
7	62.5	14.0	732	1 ACET_HUMAN	P22966 homo sapien
8	62.5	14.0	1067	1 HGBB_HAEIN	Q9kiv1 haemophilus
9	62.5	14.0	1306	1 ACE_HUMAN	P12821 homo sapien
10	61.5	13.7	299	1 LAXC_LACIA	P42096 lactococcus
11	61.5	13.7	299	1 LAXP_LACIA	P23496 lactococcus
12	61	13.6	426	1 YG43_ARCFU	O28630 archaeoglob
13	61	13.6	599	1 CENB_HUMAN	P07199 homo sapien
14	60	13.4	240	1 Y162_HELPJ	Q9zmr9 helicobacte
15	60	13.4	240	1 Y162_HELPJ	O42970 helicobacte
16	60	13.4	613	1 CGAA_CLOBI	Q45882 clostridium
17	59.5	13.3	363	1 R23A_HUMAN	P54725 homo sapien
18	59	13.2	251	1 FLGE_SALTY	P16323 salmonella
19	59	13.2	329	1 Y429_CHLTR	O84436 chlamydia t
20	59	13.2	520	1 C11A_PIG	P10612 sus scrofa
21	59	13.2	2163	1 BRR2_YEAST	P32639 saccharomyc
22	58.5	13.1	270	1 YKAA_CAEEL	P34261 caenorhabdi
23	58	12.9	297	1 Y1M4_YEAST	P40471 saccharomyc
24	58	12.9	891	1 DPOL_HELPY	P56105 helicobacte
25	58	12.9	1234	1 PIP3_HUMAN	Q01970 homo sapien
26	57.5	12.8	213	1 ATPO_BOVIN	P13621 bos taurus
27	57.5	12.8	331	1 FLGI_THEMEA	Q9xlm5 thermotoga
28	57.5	12.8	363	1 R23A_MOUSE	P54726 mus musculu
29	57.5	12.8	475	1 U2AF_HUMAN	P26368 homo sapien
30	57.5	12.8	536	1 MVIN_CHLMU	Q9pjb9 chlamydia m
31	57.5	12.8	638	1 YQGS_BACSU	P54496 bacillus su
32	57.5	12.8	875	1 AMD2_XENIA	P12890 xenopus lae
33	57.5	12.8	2292	1 POLG_EMCVB	P17593 encephalomy

RESULT 1

ID	YFL4_YEAST	STANDARD	PRT	481 AA
AC	P43616			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Hypothetical 52.9 kDa protein in SAPI55-YMR31 intergenic region.			
GN	YFR044C			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288C / AB972			
RX	MEDLINE=95400292; PubMed=7670463			
RA	Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,			
RA	Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,			
RA	Yamazaki M., Tashiro H., Eki T.;			
RT	"Analysis of the nucleotide sequence of chromosome VI from			
RT	Saccharomyces cerevisiae."			
RL	Nat. Genet. 10:261-268(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288C / AB972			
RX	MEDLINE=96287652; PubMed=8686379;			
RA	Eki T., Naitou M., Hagiwara H., Shibata T., Hanaoka F., Murakami Y.;			
RA	Sasanuma M., Tsuchiya Y., Shibata T.,			
RT	"Analysis of a 36.2 kb DNA sequence including the right telomere of			
RT	chromosome VI from Saccharomyces cerevisiae."			
RL	Yeast 12:149-167(1996).			
CC	- - SIMILARITY: STRONG, TO THE C-TERMINAL HALF OF YEAST YBR281C.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; D50617; BAA09283.1; -			
DR	EMBL; D44597; BAA08010.1; -			
DR	SGD; S0001940; YFR044C.			
DR	InterPro: IPR002933; Peptidase_M20.			
DR	Pfam; PF01546; Peptidase_M20; 1.			
KW	Hypothetical protein.			
SW	SEQUENCE 481 AA; 52871 MW; 3E53773A945F5EBC CRC64;			

Query Match 16.3%; Score 73; DB 1; Length 481;
Best Local Similarity 28.6%; Pred. No. 0.99;
Matches 28; Conservative 11; Mismatches 31; Indels 28; Gaps 4;
Qy 17 KTLILLDVSTKNPVRTV-----NENFLSL-----QLDPSIIHDG--WL-- 52

34	57.5	12.8	2292	1	POLG_EMCVD	P17594 encephalomy
35	57.5	12.8	3206	1	POLG_PSBMV	P29152 p genome po
36	57.5	12.8	4367	1	DYHC_NEUCR	P45443 neurospora
37	57	12.7	453	1	YBU5_YEAST	P38255 saccharomyc
38	57	12.7	750	1	PSAA_MESVI	Q9mur8 mesostigma
39	56.5	12.6	465	1	CIXG_HAEIN	P44458 h citxg pro
40	56.5	12.6	536	1	SVIN_CHLTR	O46378 chlamydia t
41	56.5	12.6	1042	1	SVI_BORBU	O51773 borrelia bu
42	56.5	12.6	1528	1	KEMI_YEAST	P22147 saccharomyc
43	56.5	12.6	1695	1	KFLA_MOUSE	P33173 mus musculu
44	56.5	12.6	2233	1	COAC_YEAST	Q00955 saccharomyc
45	56.5	12.6	2709	1	IP3R_BOVIN	Q9tuc34 bos taurus

ALIGNMENTS

Db 334 KTVIPAKVGFSGFSIRTVPMDSKLTSLVQKCDAKFKSLNPNKCRTELHLDGAYWSD 393
 Qy 53 ----DFLSSKRLVTLARGLSPAFLRFGGKRTDFLQFN 86
 Db 394 PFNAQFTAAKRAKTLVGVDPDFTRREGSGSIPITLTFQD 431

RESULT 2
 ID RK4_SPIOL STANDARD; PRT; 293 AA.
 AC 049937; 049938;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE 50S ribosomal protein L4, chloroplast precursor (R-protein L4).
 GN RPL4.
 OS Spinacia oleracea (Spinach).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.
 OX NCBI_TaxID=3562;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 51-63.
 RC STRAIN=CV. MELODY; TISSUE=Leaf;
 RX MEDLINE=98129805; PubMed=10874046;
 RA Trifa Y., Privat I., Gagnon J., Baeza L., Lerbs-Wache S.;
 RT "The nuclear RPL4 gene encodes a chloroplast protein that co-purifies
 RT with the 77-like transcription complex as well as plastid
 RT ribosomes.";
 RL J. Biol. Chem. 273:3980-3985(1998).
 RN [2]
 RP SEQUENCE OF 51-68; 82-86; 98-105; 131-137; 152-158; 211-224 AND
 RP 228-235.
 RC STRAIN=CV. ALVARO; TISSUE=Leaf;
 RX MEDLINE=20435798; PubMed=10874046;
 RA Yamaguchi K., Subramanian A.R.;
 RT "The plastid ribosomal proteins. Identification of all the proteins in
 RT the 50 S subunit of an organelle ribosome (chloroplast).";
 RL J. Biol. Chem. 275:28466-28482(2000).
 CC -!- FUNCTION: THIS PROTEIN BINDS DIRECTLY AND SPECIFICALLY TO 23S RRNA
 CC (BY SIMILARITY). MAY PLAY A ROLE IN PLASTID TRANSCRIPTIONAL
 CC REGULATION.
 CC -!- SUBCELLULAR LOCATION: Chloroplast.
 CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN COTYLEDON AND WEAKLY IN
 CC ROOTS.
 CC -!- SIMILARITY: BELONGS TO THE LAP FAMILY OF RIBOSOMAL PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Y14932; CAA75149.1; -;
 CC EMBL; X93160; CAA63651.1; -;
 CC InterPro; IPR002136; Ribosomal_L4/L1E.
 CC Pfam; PF00573; Ribosomal_L4; 1.
 CC Ribosomal protein; rRNA-binding; Chloroplast; Transit peptide.
 KW TRANSIT 1 50
 FT CHAIN 51 293 50S RIBOSOMAL PROTEIN L4.
 FT DOMAIN 265 283 ASP/GLU-RICH (ACIDIC).
 FT CONFLICT 274 274 E -> G (IN REF. 1; CAA63651).
 SQ SEQUENCE 293 AA; 32434 MW; 875C425F090826E5 CRC64;

Query Match 15.1%; Score 67.5; DB 1; Length 293;
 Best Local Similarity 29.1%; Pred. No. 2.3;
 Matches 23; Conservative 13; Mismatches 20; Indels 23; Gaps 3;
 Qy 5 PLPVDRAAGLKEKTLILDVSTKNPVRTVNEFNLSQLDLP-----SIHGDWDLFLSSKR 59

Db 54 PLPLNLSGSK-----VGSETFLNLTAPPEKARAVVHRLGLTHLQNK 96
 Qy 60 LVTLARGLSPAFLRFGGKR 78
 Db 97 RGT-ASTLTRAERVGGRK 114

RESULT 3
 ID CAT2_MAIZE STANDARD; PRT; 491 AA.
 AC P12365;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Catalase isozyme 2 (EC 1.11.1.6).
 GN CAT2.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoideae; Andropogoneae; Zea.
 OX NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. W64A;
 RA Guan L., Ruzsa S., Skadsen R.W., Scandalios J.G.;
 RT "Comparison of the cat2 complementary DNA sequences of a normal
 RT catalase activity line (W64A) and a high catalase activity line (R6-
 RT 67) of maize.";
 RL Plant Physiol. 96:1379-1381(1991).
 RN [2]
 RP PRELIMINARY SEQUENCE FROM N.A.
 RC STRAIN=R6-67;
 RX MEDLINE=98016183; PubMed=2821546;
 RA Becharads L.A., Skadsen R.W., Scandalios J.G.;
 RT "Isolation and characterization of a cDNA clone for the Cat2 gene in
 RT maize and its homology with other catalases.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:6830-6834(1987).
 RN [3]
 RP REVISIONS TO C-TERMINUS.
 RX MEDLINE=90370897; PubMed=2395887;
 RA Becharads L.A., Skadsen R.W., Scandalios J.G.;
 RT "Isolation and characterization of a cDNA clone for the Cat2 gene in
 RT maize and its homology with other catalases.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:6927-6927(1990).
 CC -!- FUNCTION: DECOMPOSES HYDROGEN PEROXIDE IN WATER AND OXYGEN; SERVES
 CC TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE.
 CC -!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
 CC -!- Cofactor: HEME GROUP.
 CC -!- SUBUNIT: HOMOTETRAMER.
 CC -!- SUBCELLULAR LOCATION: PEROXISOMAL OR CYTOPLASMIC.
 CC -!- SIMILARITY: BELONGS TO THE CATALASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X54819; CAA38588.1; -;
 CC EMBL; J02976; AAA33440.1; -;
 CC PIR; A36062; A36062.
 CC PIR; S18819; S18819.
 CC HSP; P00432; 4BLC.
 CC MaizeDB; 13855;
 CC InterPro; IPR002226; Catalase.
 CC Pfam; PF00199; catalase; 1.
 CC PRINTS; PR00067; CATALASE.
 CC ProDom; PD000510; CATALASE.
 CC PROSITE; PS00437; CATALASE_1; 1.
 CC PROSITE; PS00438; CATALASE_2; 1.

RESULT 6		UL87_EBV		STANDARD;		PRT;		575 AA.	
AC	P25215;	01-MAY-1992	(Rel. 22, Created)						
DT	01-MAY-1992	(Rel. 22, Last sequence update)							
DT	01-OCT-1996	(Rel. 34, Last annotation update)							
DE	Protein B(C)RFL								
GN	B(C)RFL								
OS	Epstein-barr virus (strain B95-8) (Human herpesvirus 4).								
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;								
OC	Gammarherpesvirinae; Lymphocryptovirus.								
OX	NCBI_TaxID=10377;								
RI	[1]								
RN	SEQUENCE FROM N.A.								
RP	MEDLINE=94270667; PubMed=6087149;								
RA	Beer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,								
RA	Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,								
RA	Tufnell P.S., Barrell B.G.,								
RT	"DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";								
RL	Nature 310:207-211(1984).								
CC	-1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS TOGETHER EBV BCRFL,								
CC	HSV-6 U58, HVS-1 24 AND HCMV UL87.								
CC	-1- CAUTION: THERE ARE 2 BCRFL GENES. ONE IS WRITTEN WITH A LOWER								
CC	CASE C (B(C)RFL) AND THE OTHER WITH A CAPITAL C. (BCRFL).								
CC	-----								
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration								
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -								
CC	the European Bioinformatics Institute. There are no restrictions on its								
CC	use by non-profit institutions as long as its content is in no way								
CC	modified and this statement is not removed. Usage by and for commercial								
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/								
CC	or send an email to license@isb-sib.ch).								
CC	-----								
DR	EMBL; V01555; CAA24795.1; ALT_INIT.								
DR	InterPro; IPR004285; UL87.								
DR	Pfam; PF03043; UL87; 1.								
SQ	SEQUENCE 575 AA; 64304 MW; 79B3206D8DEB7337 CRC64;								
Query Match 14.0%; Score 62.5; DB 1; Length 575;									
Best Local Similarity 27.8%; Pred. NO. 19;									
Matches 20; Conservative 16; Mismatches 19; Indels 17; Gaps 4;									
QY	5	PLPVDRAA---GLKETLLLDVSTKNPVTYVNFLS-----LQDPSLIH-----D 49							
DB	406	PVTHREAKISKIKRNLFTLLELRNRSQIQVHKRFLGGLDCASLLRLDPSCINRIASE 465							
QY	50	GWLDFLSSKRLV 61							
DB	466	GLDFD---SKRSI 475							
RESULT 7									
AC	ACET HUMAN								
ID	ACET_HUMAN								
AC	P22966;								
DT	01-AUG-1991 (Rel. 19, Created)								
DT	01-AUG-1991 (Rel. 19, Last sequence update)								
DT	01-MAR-2002 (Rel. 41, Last annotation update)								
DE	Angiotensin-converting enzyme, testis-specific isoform precursor								
DE	(EC 3.4.15.1) (ACE-T) (Dipeptidyl carboxypeptidase I) (Kininase II).								
GN	DCP1 OR DCP OR ACE.								
OS	Homo sapiens (Human).								
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;								
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.								
OX	NCBI_TaxID=9606;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RX	MEDLINE=90046671; PubMed=2554286;								
RA	Ehlers M.R.W., Fox E.A., Strydom D.J., Riordan J.F.;								

"Molecular cloning of human testicular angiotensin-converting enzyme: the testis isozyme is identical to the C-terminal half of endothelial angiotensin-converting enzyme.";

Proc. Natl. Acad. Sci. U.S.A. 86:7741-7745(1989).

[2]

SEQUENCE FROM N.A.

MEDLINE=89338720; PubMed=2547653;

Lattion A.L., Soubrier F., Allegrini J., Hubert C., Corvol P.,

Alhenc-Gelas F.; transcript of the angiotensin I-converting enzyme

"The testicular transcript of the angiotensin I-converting enzyme

encodes for the ancestral, non-duplicated form of the enzyme.";

FEBS Lett. 252:99-104(1989).

[3]

SEQUENCE FROM N.A., AND VARIANTS P-32; G-49 AND S-712.

MEDLINE=99251580; PubMed=10319862;

Rieder M.J., Taylor S.L., Clark A.G., Nickerson D.A.;

"Sequence variation in the human angiotensin converting enzyme.";

Nat. Genet. 22:59-62(1999).

[4]

ZINC-BINDING.

MEDLINE=91308093; PubMed=1649623;

Ehlers M.R., Riordan J.F.;

"Angiotensin-converting enzyme: zinc- and inhibitor-binding

stoichiometries of the somatic and testis isozymes.";

Biochemistry 30:7118-7126(1991).

-1- FUNCTION: CONVERTS ANGIOTENSIN I TO ANGIOTENSIN II BY RELEASE OF THE

TERMINAL HIS-LEU, THIS RESULTS IN AN INCREASE OF THE

VASOCONSTRICTOR ACTIVITY OF ANGIOTENSIN.

-1- CATALYTIC ACTIVITY: Release of a C-terminal dipeptide,

oligopeptide-l-xaa-xbb, when xaa is not pro, and xbb is neither

asp nor glu. Converts angiotensin I to angiotensin II.

-1- COFACTOR: BINDS 1 ZINC ION.

-1- SUBCELLULAR LOCATION: Type I membrane protein.

-1- ALTERNATIVE PRODUCTS: THE TESTICULAR ANGIOTENSIN-CONVERTING ENZYME

IS TRANSCRIBED FROM THE SAME GENE AS THE SOMATIC ISOFORM, PROBABLY

FROM AN ALTERNATIVE START SITE.

-1- TISSUE SPECIFICITY: SPERMATOCYTES, ADULT TESTIS.

-1- INDUCTION: EXPRESSION IS THOUGHT TO BE SUBJECT TO HORMONAL

REGULATION BY ANDROGENS.

-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M2.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL Outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL; M26657; AAA60611.1; -;

EMBL; X16295; CAA34362.1; -;

EMBL; AF118569; AAD28561.1; -;

PIR; S05238; S05238.

PIR; A33979; A33979.

MEROPS; M02.004; -;

MIM; 106180; -;

InterPro; IPR001548; Peptidase_M2.

InterPro; IPR00130; zn_Mtpeptdse.

Pfam; PF01401; Peptidase_M2; 1.

PRINTS; PR00791; PEPDIPTASEA.

PRODOM; PD004184; Peptidase_M2; 1.

PROSITE; PS00142; ZINC_PROTEASE; 1.

KW Hydrolase; Metalloprotease; Carboxypeptidase; zinc; Dipeptidase;

KW Glycoprotein; Transmembrane; Testis; Signal; Alternative splicing;

KW Polymorphism.

FT SIGNAL 1 31

FT CHAIN 32 732

FT

FT ANGIOTENSIN-CONVERTING ENZYME,

FT TESTIS-SPECIFIC ISOFORM.

FT EXTRACELLULAR (POTENTIAL).

FT

FT DOMAIN 32 684

FT TRANSMEM 685 701

FT POTENTIAL.

FT CYTOPLASMIC (POTENTIAL).

FT DOMAIN 702 732

FT METAL 414 414

FT ZINC (CATALYTIC) (BY SIMILARITY).

FT ACT_SITE 415 415

FT BY SIMILARITY.


```
FT METAL 418 418 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 442 442 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 457 457 POTENTIAL.
FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 617 617 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 32 32 S -> P.
FT VARIANT 49 49 /FTIG=VAR_011710.
FT VARIANT 712 712 S -> G.
FT VARIANT 732 732 /FTIG=VAR_011711.
FT SEQUENCE 732 AA; 83330 MW; 80E0D19CFA642313 CRC64;

Query Match 14.0%; Score 62.5; DB 1; Length 732;
Best Local Similarity 32.4%; Pred. No. 26;
Matches 22; Conservative 8; Mismatches 19; Indels 19; Gaps 3;

QY 2 DRRPLPVDRAAGLKETLLIDVSTKNPRTV-NENFLSLQLDPSIIH----- 48
Db 154 ERAALP---AOELEENKILLDMETTSVATVCHPGNSCLQLEPDLTNVMTSRKYEDLL 210
QY 49 ---DGLWD 53
Db 211 WAVEGWRD 218

RESULT 8
HGBB_HAEBIN STANDARD; PRT; 1067 AA.
AC Q9KIV1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemoglobin and hemoglobin-haptoglobin binding protein B precursor
DE (Hemoglobin binding protein B).
GN HGBB.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTIH N182;
RX MEDLINE=20316037; PubMed=10858226;
RA Cope L.D., Hrkal Z., Hansen E.J.;
RT "Detection of phase variation in expression of proteins involved in
hemoglobin and hemoglobin-haptoglobin binding by nontypeable
Haemophilus influenzae."
RL Infect. Immun. 68:4092-4101(2000).
CC -!- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE
HEMOGLOBIN/HAPTOGLOBIN COMPLEX OF THE HUMAN HOST AND IS REQUIRED
FOR HEME UPTAKE.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCAA
REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAGING.
ADDITION OR LOSS OF CCAA REPEAT UNITS WOULD CHANGE THE READING
FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A
WAY TO AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST.
CC -!- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR FAMILY;
HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
```

```
CC entties requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; AF221059; AAF80177.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC.1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Transport; TonB box; Multigene family; Signal;
KW Receptor; Repeat.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 1067 HEMOGLOBIN AND HEMOGLOBIN-HAPTOGLOBIN
BINDING PROTEIN B.
FT DOMAIN 26 49 6 X 4 AA TANDEM REPEATS OF Q-P-T-N.
FT REPEAT 26 29 1.
FT REPEAT 30 33 2.
FT REPEAT 34 37 3.
FT REPEAT 38 41 4.
FT REPEAT 42 45 5.
FT REPEAT 46 49 6.
FT SITE 59 66 TONB BOX.
FT SITE 1050 1067 TONB C-TERMINAL BOX.
FT SEQUENCE 1067 AA; 122471 MW; 29D295DC4747632E CRC64;

Query Match 14.0%; Score 62.5; DB 1; Length 1067;
Best Local Similarity 27.8%; Pred. No. 40;
Matches 27; Conservative 18; Mismatches 33; Indels 19; Gaps 5;

QY 1 GDRRLPVDRAAGLKETLI-----LLDVSTKNPVTNEN-----FLSLQLDPSI 46
Db 598 GEKKRPDISVAGCANGTLLSHDIGDKTYLIPVTKNNVLYFGDNVOLTSLWGLDLNRY 657
QY 47 IHGWLDFLSKRLVTLARGL-SPAFLRFGGKRTDFL 82
Db 658 DHVKYLP--SYDKNIPVPKGLITGLFKRF--KSTDIV 690

RESULT 9
ACE_HUMAN STANDARD; PRT; 1306 AA.
ID ACE_HUMAN
AC P12821;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Angiotensin-converting enzyme, somatic isoform precursor (EC 3.4.15.1)
DE (ACE) (Dipeptidyl carboxypeptidase I) (Kinase II) (CD143 antigen).
GN DCP1 OR DCP OR ACE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89071703; PubMed=2849100;
RA Soubrier F., Alhenc-Gelas F., Hubert C., Allegrini J., John M.,
RA Tregear G., Corbol P.;
RT "Two putative active centers in human angiotensin I-converting enzyme
revealed by molecular cloning."
RL Proc. Natl. Acad. Sci. U.S.A. 85:9386-9390(1988).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANTS S-261; W-561 AND S-1286.
RX MEDLINE=99251580; PubMed=10319862;
RA Rieder M.J., Taylor S.L., Clark A.G., Nickerson D.A.;
RT "Sequence variation in the human angiotensin converting enzyme."
RL Nat. Genet. 22:59-62(1999).
RN [3]
RP PARTIAL SEQUENCE OF 30-46.
RC TISSUE=Lung.
RX MEDLINE=90110025; PubMed=2558109;
RA Takeuchi K., Shimizu T., Ohishi N., Seyama Y., Takaku F.,
RA Totsumoto H.;
RT "Purification of human lung angiotensin-converting enzyme by high-
```



```

QY 79 TDFLQFQNL 87
   || | ||
Db 236 VDFDFPNL 244

RESULT 11
LAXP_LACLA
ID LAXP_LACLA STANDARD; PRT; 299 AA.
AC P23496;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Lax protein, plasmid.
GN LACX.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OG Plasmid pmg820.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91093107; PubMed=2125052;
RA de Vos W.M., Boerrigter I., van Rooyen R.J., Reiche B.,
RA Hengstenberg W.;
RT "Characterization of the lactose-specific enzymes of the
RT phosphotransferase system in Lactococcus lactis.";
RL J. Biol. Chem. 265:22554-22560(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M60447; AAA25184.1; -.
DR PIR; D23696; D23696.
DR InterPro; IPR003843; UPF0010_DUF.
DR Pfam; PF02693; UPF0010; 1.
KW Plasmid.
SQ SEQUENCE 299 AA; 34519 MW; 715599C040DCD608 CRC64;

Query Match 13.7%; Score 61.5; DB 1; Length 299;
Best Local Similarity 42.0%; Pred. No. 11;
Matches 29; Conservative 3; Mismatches 26; Indels 11; Gaps 5;

QY 21 LLDVSTKNPVRTNENFLSLQDPSII-HDG-WLDFLSSKRLVTLARGLSPAFLRFGGR 78
   ||| : | || || | : || || || | || |
Db 185 LLDLQDRTPL---ENOKSLDLDYSLFSDAITLDRKS-RSVTLRSRKS GK-----GLR 235

QY 79 TDFLQFQNL 87
   || | ||
Db 236 VDFDFPNL 244

RESULT 12
YG43_ARCFU
ID YG43_ARCFU STANDARD; PRT; 426 AA.
AC O28630;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF1643.
GN AF1643.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyripides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -!- SIMILARITY: BELONGS TO THE UBID FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000989; AAB89599.1; -.
DR TIGR; AF1643; -.
DR InterPro; IPR002830; UPF0096.
DR Pfam; PF01977; UPF0096; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 426 AA; 47715 MW; 7EBCE276F501E2D8 CRC64;

Query Match 13.6%; Score 61; DB 1; Length 426;
Best Local Similarity 29.9%; Pred. No. 20;
Matches 23; Conservative 7; Mismatches 21; Indels 26; Gaps 4;

QY 13 GLK-EKTLILID-----VSTKNPVRTNENFLSLQDPSIIHDGWLDFLSSKRLV 61
   ||| : | || || | : || || || | || |
Db 2 GIKTKNFICLKRTFSMNLRAISVANPV-----QLEEEIKHDEVVFLSKSNLL 51

QY 62 TLARGLSPAFLRFGGR 78
   | | ||
Db 52 D-----KPVILNVEGK 63

RESULT 13
CENB_HUMAN
ID CENB_HUMAN STANDARD; PRT; 599 AA.
AC P07139;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Major centromere autoantigen B (Centromere protein B) (CENP-B).
GN CENPB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91372020; PubMed=1893793;
RA Sullivan K.F., Glass C.A.;
RT "CENP-B is a highly conserved mammalian centromere protein with
RT homology to the helix-loop-helix family of proteins.";
RL Chromosoma 100:360-370(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,

```

RA Buck D., Burrill W., Butler A.P., Carder C., Carter N.P., Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M., Clegg S., Cobley V.E., Collier R.E., Connor R., Corby N.R., Coulson A., Coville G.J., Deadman R., Dhani P., Dunn M., Ellington A.G., Frankland J.A., Fraser A., French L., Garner P., Grahm D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E., Hammond S., Hartley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J., Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D., Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S., Leivasalho M.H., Leversha M., Lloyd C., Lloyd D.M., Lovell J.D., Marsh V.L., Martin S.L., McConnachie L.J., Mcclay K., McMurray A.A., Milne S., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T., Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I., Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H., Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showken R., Sims S., Suce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E., Swann M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A., Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A., Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S., Rogers J.;
 RL "The DNA sequence and comparative analysis of human chromosome 20.";
 RT Nature 414:865-871(2001).
 RN [3]
 RP SEQUENCE OF 6-599 FROM N.A.
 RX MEDLINE=97166180; PubMed=2435739;
 RA Earnshaw W.C., Sullivan K.F., Machlin P.S., Cooke C.A., Kaiser D.A., Pollard T.D., Rothfield N.F., Cleveland D.W.;
 RT "Molecular cloning of cDNA for CENP-B, the major human centromere autoantigen.";
 RN J. Cell Biol. 119:1413-1427(1992).
 RN [4]
 RP J. Cell Biol. 104:817-829(1987).
 RN [4]
 RP SUBUNITS AND DOMAINS.
 RX MEDLINE=93107144; PubMed=1469042;
 RA Yoda K., Kitagawa K., Masumoto H., Muro Y., Okazaki T.;
 RT "A human centromere protein, CENP-B, has a DNA binding domain containing four potential alpha helices at the NH2 terminus, which is separable from dimerizing activity.";
 RN J. Cell Biol. 119:1413-1427(1992).
 RN [5]
 RP STRUCTURE BY NMR OF 1-56.
 RX MEDLINE=98119825; PubMed=9451007;
 RA Iwahara J., Kitagawa T., Kitagawa K., Masumoto H., Okazaki T., Yokoyama S.;
 RT "A helix-turn-helix structure unit in human centromere protein B (CENP-B).";
 RN EMBO J. 17:827-837(1998).
 CC -1- FUNCTION: INTERACTS WITH CENTROMERIC HETEROCHROMATIN IN CHROMOSOMES AND BINDS TO A SPECIFIC SUBSET OF ALPHOID SATELLITE DNA, CALLED THE CENP-B BOX. MAY ORGANIZE ARRAYS OF CENTROMERE SATELLITE DNA INTO A HIGHER ORDER STRUCTURE WHICH THEN DIRECTS CENTROMERE FORMATION AND KINETOCORE ASSEMBLY IN MAMMALIAN CHROMOSOMES.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: STRONG, WITH OTHER MAMMALIAN CENP-B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X55039; CAA38879.1; -
 CC EMBL: AL109804; CAC17547.1; -
 CC EMBL: X05299; CAA28918.1; -
 CC PIR: A27272; A27272;
 CC PIR: S18735; S18735;
 CC PDB: 1BW6; 07-OCT-98.
 CC MIM: 117140; -
 CC Chromosomal protein; Nuclear protein; DNA-binding; Centromere;

KW 3D-structure. 1 125
 FT DNA_BIND 404 465 GLU-RICH (ACIDIC).
 FT DOMAIN 508 538 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 583 593 R -> M (IN REF. 3).
 FT CONFLICT 592 593 VR -> LL (IN REF. 3).
 SQ SEQUENCE 599 AA; 65171 MW; 9B4B7DB957A914AA CRC64;
 Query Match 13.6%; Score 61; DB 1; Length 599;
 Best Local Similarity 31.4%; Pred. No. 30;
 Matches 22; Conservative 8; Mismatches 32; Indels 8; Gaps 2;
 QY 10 RAAGLKEKTLILDVSTKNPVRTVNENFLSLQDPSIIHDGWLDFSSKRLVTLARGLSP 69
 Db 89 RAAGLPVKGIIL---KEKALRTAE---LGMDDFTASNGWLDRFRRRHGVVSCSGVAR 140
 QY 70 AFLRFGGKRT 79
 Db 141 ARARNAAPRT 150
 RESULT 14
 Y162_HELPJ
 ID Y162_HELPJ STANDARD; PRT; 240 AA.
 AC Q9ZMR9;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein JHP0149.
 GN JHP0149.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 OX NCBI_TaxID=85963;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C., Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G., Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C., Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F., Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori.";
 RN Nature 397:176-180(1999).
 CC -1- SIMILARITY: BELONGS TO THE UPF0082 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE001453; AAD05730.1; -
 CC InterPro: IPR002876; DUF28.
 CC Pfam: PF01709; DUF28; 1.
 CC ProDom: PD004323; DUF28; 1.
 CC Hypothetical protein; Complete proteome.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 240 AA; 27040 MW; EC304F29A8E3262C CRC64;
 Query Match 13.4%; Score 60; DB 1; Length 240;
 Best Local Similarity 28.9%; Pred. No. 13;
 Matches 13; Conservative 15; Mismatches 15; Indels 2; Gaps 1;
 QY 19 LILDVSTKNPVRTVN---ENFLSLQDPSIIHDGWLDFSSKRLV 61
 Db 92 LIIMECTDNPRTIANLKSIFNKQTGASIVPNSGLEMFENRKS 136

```
RESULT 15
Y162_HELPY
ID Y162_HELPY STANDARD; PRT; 240 AA.
AC O24970;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein HP0162.
GN HP0162.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G., B.A.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -!- SIMILARITY: BELONGS TO THE UPF0082 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000537; AAD07231.1; -.
CC TIGR; HP0162; -.
DR InterPro: IPR002876; DUF28.
DR Pfam; PF01709; DUF28; 1.
DR ProDom; PD004323; DUF28; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 240 AA; 27127 MW; 6BB7E0DC87B35F2E CRC64;

Query Match 13.4%; Score 60; DB 1; Length 240;
Best Local Similarity 28.9%; Pred. No. 13;
Matches 13; Conservative 15; Mismatches 15; Indels 2; Gaps 1;

QY 19 LILLDVSTKNPVRTVN--ENFLSLQLDPSIIHDGWLDFLSKRLV 61
|:::| | | |:::| | | |:::|
DB 92 LIIMECHTDNPTRIANLKSIFYNKTQASIVPNGSLEFMNRKSV 136
```

Search completed: July. 30, 2002, 08:31:27
Job time: 974 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:30:41 ; Search time 123.94 Seconds
(without alignments)
122.830 Million cell updates/sec

Title: US-09-836-461-2_COPY_42_129
Perfect score: 448
Sequence: 1 GDRPLPVDRAAGLKEKTLI.....PAFLRFGGKRTDFLQFQNL 88

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL19:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvrius:*
 - 16: sp_bacteriap:*
 - 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	448	100.0	480	4 Q9HB39	Q9hb39 homo sapien
2	448	100.0	534	4 Q9HB38	Q9hb38 homo sapien
3	448	100.0	592	4 Q9HB37	Q9hb37 homo sapien
4	155	34.6	536	11 Q9QZP8	Q9qzf8 rattus norv
5	151.5	33.8	545	4 Q9UL39	Q9ul39 homo sapien
6	150.5	33.6	543	4 Q9V251	Q9y251 homo sapien
7	140.5	31.4	545	6 Q9MYI0	Q9myi0 bos taurus
8	132.5	29.6	523	13 Q9QYK5	Q9qyk5 gallus gall
9	72.5	16.2	304	10 Q9C5M5	Q9c5m5 arabidopsis
10	72.5	16.2	334	10 Q9LFY5	Q9lfy5 arabidopsis
11	70.5	15.7	527	10 Q9LRC8	Q9lrc8 scutellaria
12	68.5	15.3	329	10 Q9FG13	Q9fg13 arabidopsis
13	68.5	15.3	694	16 Q98PE1	Q98pe1 mycoplasma
14	66	14.7	305	12 Q9LMM5	Q9lmm5 lumpy skin
15	65.5	14.6	1391	5 Q17772	Q17772 caenorhabdi
16	65	14.5	921	2 Q9EYG5	Q9eyg5 actinobacil

17	65	14.5	2288	3	O60033	O60033 emericeila
18	64.5	14.4	199	5	Q9BI00	Q9bi00 globodera p
19	64.5	14.4	702	10	Q9FLL6	Q9fil6 arabidopsis
20	64.5	14.4	881	5	Q9GR11	Q9gr11 paramecium
21	64.5	14.4	2126	2	Q93LR2	Q93lr2 microcystis
22	64.5	14.4	2126	2	Q93LR1	Q93lr1 microcystis
23	64	14.3	369	10	Q42606	Q42606 arabidopsis
24	64	14.3	536	10	Q9C825	Q9c825 arabidopsis
25	64	14.3	1477	3	O74218	O74218 candida alb
26	63.5	14.2	497	5	Q18652	Q18652 caenorhabdi
27	63.5	14.2	756	16	Q988J0	Q988j0 rhizobium 1
28	63.5	14.2	935	5	Q9VE79	Q9ve79 drosophila
29	63.5	14.2	2404	5	Q9VE34	Q9ve34 drosophila
30	62.5	14.0	326	16	Q97DT6	Q97dt6 clostridium
31	62.5	14.0	369	10	Q96532	Q96532 arabidopsis
32	62.5	14.0	694	4	Q15540	Q15540 homo sapien
33	62.5	14.0	732	6	Q9GLN6	Q9gln6 pan troglod
34	62.5	14.0	1304	6	Q9GLN7	Q9gln7 pan troglod
35	62.5	14.0	2126	2	Q9SIA8	Q9sls8 microcystis
36	62	13.8	144	6	Q9TU38	Q9tu38 oryctolagus
37	62	13.8	292	10	Q9LG45	Q9lg45 oryza sativ
38	62	13.8	295	16	Q98H10	Q98h10 rhizobium 1
39	62	13.8	749	4	O60741	O60741 homo sapien
40	62	13.8	822	6	Q9MZS1	Q9mzs1 oryctolagus
41	62	13.8	830	16	Q9KGI3	Q9kgi3 bacillus ha
42	62	13.8	910	11	O54899	O54899 mus musculus
43	62	13.8	910	11	Q9UKB0	Q9jkb0 rattus norv
44	62	13.8	910	11	O88704	O88704 mus musculus
45	61.5	13.7	126	17	O57867	O57867 pyrococcus

ALIGNMENTS

RESULT 1

Q9HB39 PRELIMINARY; PRT; 480 AA.

AC Q9HB39; 01-MAR-2001 (Tremblrel. 16, Created)

DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)

DE HEPARANASE-LIKE PROTEIN HPA2A.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=20483645; PubMed=11027606;

RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,

RA Hircok M., Patel S., Barry E., Stubbfield C., Terrett J., Page M.;

RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian

RT Heparanase Family Member.";

RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).

DR EMBL; AF282885; AAC23421.1; -

SQ SEQUENCE 480 AA; 53900 MW; F75F89F67AC1FF83 CRC64;

Query Match 100.0%; Score 448; DB 4; Length 480;
Best Local Similarity 100.0%; Pred. No. 3.5e-43;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GDRPLPVDRAAGLKEKTLIILLDVSTKNPVTNENFLSLQDPSIIHDGWLDFLSKRL 60
|||||

Db 42 GDRPLPVDRAAGLKEKTLIILLDVSTKNPVTNENFLSLQDPSIIHDGWLDFLSKRL 101
|||||

Qy 61 VTLAGLSPAFLRFGGKRTDFLQFQNL 88
|||||

Db 102 VTLAGLSPAFLRFGGKRTDFLQFQNL 129
|||||

RESULT 2

Q9HB38

ID Q9HB38 PRELIMINARY; PRT; 534 AA.
AC Q9HB38;
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
DE HEPARANASE-LIKE PROTEIN HPA2B.
DE HEPARANASE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20483645; PubMed=11027606;
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,
RA Hircok M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.,
RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian
RT Heparanase Family Member.";
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).
DR EMBL; AF282886; AAG23422.1; -.
SQ SEQUENCE 534 AA; 60063 MW; C3DE5E900CB338C4 CRC64; 0

Query Match 100.0%; Score 448; DB 4; Length 534;
Best Local Similarity 100.0%; Pred. No. 3.9e-43;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GDRPLPVDRAAGLKEKTLILLDVSTKNPVTNENFLSLQLDPSIHDGWLDFLSSKRL 60
DB 42 GDRPLPVDRAAGLKEKTLILLDVSTKNPVTNENFLSLQLDPSIHDGWLDFLSSKRL 101
QY 61 VTLARGLSPAFLRGGKRTDFLOFNLR 88
DB 102 VTLARGLSPAFLRGGKRTDFLOFNLR 129

RESULT 3
Q9HB37
ID Q9HB37 PRELIMINARY; PRT; 592 AA.
AC Q9HB37;
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
DE HEPARANASE-LIKE PROTEIN HPA2C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20483645; PubMed=11027606;
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,
RA Hircok M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.,
RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian
RT Heparanase Family Member.";
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).
DR EMBL; AF282887; AAG23423.1; -.
SQ SEQUENCE 592 AA; 66580 MW; 95C384AD9A74258E CRC64;

Query Match 100.0%; Score 448; DB 4; Length 592;
Best Local Similarity 100.0%; Pred. No. 4.5e-43;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GDRPLPVDRAAGLKEKTLILLDVSTKNPVTNENFLSLQLDPSIHDGWLDFLSSKRL 60
DB 42 GDRPLPVDRAAGLKEKTLILLDVSTKNPVTNENFLSLQLDPSIHDGWLDFLSSKRL 101
QY 61 VTLARGLSPAFLRGGKRTDFLOFNLR 88
DB 102 VTLARGLSPAFLRGGKRTDFLOFNLR 129

RESULT 4

Q9QZF8
ID Q9QZF8 PRELIMINARY; PRT; 536 AA.
AC Q9QZF8;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE HEPARANASE.
GN HEP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RA Podyna K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.,
RT "Heparanase from parathyroid cell line.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF184967; AAF04563.1; -.
SQ SEQUENCE 536 AA; 60569 MW; 6208B1FD9EE28421 CRC64;

Query Match 34.6%; Score 155; DB 11; Length 536;
Best Local Similarity 46.3%; Pred. No. 2e-09;
Matches 40; Conservative 13; Mismatches 31; Indels 2; Gaps 2;
QY 1 GDRPLPVDRAAGLKEKTLILLDVSTKNPVTNENFLSLQLDPSIHDGWLDFLSSKRL 58
DB 12 GRURALTQGTAGTAPTKDVVDLEFYTKRLQFQSVPSFLSTIDASLATDPRFLTLGSP 71
QY 59 RLVTARGLSPAFLRGGKRTDFLOF 84
DB 72 RLVTARGLSPAFLRGGKRTDFLOF 97

RESULT 5
Q9UL39
ID Q9UL39 PRELIMINARY; PRT; 545 AA.
AC Q9UL39;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE HEPARANASE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=PLACENTA;
RX MEDLINE=20229546; PubMed=10764835;
RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.,
RT "Heparanase expression in invasive trophoblasts and acute vascular
RT damage.";
RL Glycobiology 10:467-475(2000).
DR EMBL; AF084467; AAD54516.1; -.
SQ SEQUENCE 545 AA; 61418 MW; 67B80ACD73C5A9A1 CRC64;

Query Match 33.8%; Score 151.5; DB 4; Length 545;
Best Local Similarity 47.1%; Pred. No. 5.2e-09;
Matches 33; Conservative 16; Mismatches 20; Indels 1; Gaps 1;
QY 16 EKTLLLDVSTKNPVTNENFLSLQLDPSIHDGWLDFLSSKRLVTLARGLSPAFLR 74
DB 37 QQDVVDLDFTEQEPHLVSPFSLVTIDANLATDPRFLILGSPKRLTARGLSPAYLR 96
QY 75 GGRKRTDFLOF 84
DB 97 GGRKRTDFLOF 106

RESULT 6
Q9Y251

ID Q9Y251 PRELIMINARY; PRT; 543 AA.
AC Q9Y251;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE HEPARANASE.
GN HPA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=99321249; PubMed=10395326;
RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,
RA Parish C.R.;
RT "Cloning of mammalian heparanase, an important enzyme in tumor
RT invasion and metastasis";
RL Nat. Med. 5:803-809(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Vlodavsky I., Friedman Y., Elkin M., Aingorn H., Atzmon R.,
RA Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I.,
RA Spector L., Pecker I.;
RT "Mammalian heparanase: a novel gene involved in tumor progression and
RT metastasis";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99377052; PubMed=10446189;
RA Toyoshima M., Nakajima M.;
RT "Human heparanase. Purification, characterization, cloning, and
RT expression";
RL J. Biol. Chem. 274:24153-24160(1999).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=9935379; PubMed=10405343;
RA Kussie P.H., Hulmes J.D., Ludwig D., Patel S., Navarro E.C.,
RA Seddon A.P., Giorgio N.A., Bohlen P.;
RT "Cloning and Functional Expression of a Human Heparanase Gene";
RL Biochem. Biophys. Res. Commun. 261:183-187(1999).
DR EMBL; AF165154; AAD45379.1; -
DR EMBL; AF144325; AAD41342.1; -
DR EMBL; AF155510; AAD54941.1; -
DR EMBL; AF152376; AAD45669.1; -
SQ SEQUENCE 543 AA; 61176 MW; AD262EC267334AB2 CRC64;

Query Match 33.6%; Score 150.5; DB 4; Length 543;
Best Local Similarity 43.5%; Pred. No. 6.7e-09;
Matches 37; Conservative 17; Mismatches 28; Indels 3; Gaps 2;

QY 1 GDRPLPVDRAAGLKEKTLILLDVSTKNPVTNENFLSLQDPSIHD-GWLDPLSKRLVTLAR 59
Db 22 GPLSPGALPRA--QAQDVVDLDFFTQEPHLVSPFSLVTIDANLATDPRFLILGSPK 79
QY 60 LVTLARGLSPAFRLFGGKRTDFLOF 84
Db 80 LRTLARGLSPAYLRFGGTKTDFLIF 104

RESULT 7
Q9MYO0 PRELIMINARY; PRT; 545 AA.
AC Q9MYO0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HEPARANASE.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

ID Q9Y251 PRELIMINARY; PRT; 543 AA.
AC Q9Y251;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE HEPARANASE.
GN HPA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=99321249; PubMed=10395326;
RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,
RA Parish C.R.;
RT "Cloning of mammalian heparanase, an important enzyme in tumor
RT invasion and metastasis";
RL Nat. Med. 5:803-809(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Vlodavsky I., Friedman Y., Elkin M., Aingorn H., Atzmon R.,
RA Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I.,
RA Spector L., Pecker I.;
RT "Mammalian heparanase: a novel gene involved in tumor progression and
RT metastasis";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99377052; PubMed=10446189;
RA Toyoshima M., Nakajima M.;
RT "Human heparanase. Purification, characterization, cloning, and
RT expression";
RL J. Biol. Chem. 274:24153-24160(1999).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=9935379; PubMed=10405343;
RA Kussie P.H., Hulmes J.D., Ludwig D., Patel S., Navarro E.C.,
RA Seddon A.P., Giorgio N.A., Bohlen P.;
RT "Cloning and Functional Expression of a Human Heparanase Gene";
RL Biochem. Biophys. Res. Commun. 261:183-187(1999).
DR EMBL; AF165154; AAD45379.1; -
DR EMBL; AF144325; AAD41342.1; -
DR EMBL; AF155510; AAD54941.1; -
DR EMBL; AF152376; AAD45669.1; -
SQ SEQUENCE 543 AA; 61176 MW; AD262EC267334AB2 CRC64;

Query Match 33.6%; Score 150.5; DB 4; Length 543;
Best Local Similarity 43.5%; Pred. No. 6.7e-09;
Matches 37; Conservative 17; Mismatches 28; Indels 3; Gaps 2;

QY 1 GDRPLPVDRAAGLKEKTLILLDVSTKNPVTNENFLSLQDPSIHD-GWLDPLSKRLVTLAR 59
Db 22 GPLSPGALPRA--QAQDVVDLDFFTQEPHLVSPFSLVTIDANLATDPRFLILGSPK 79
QY 60 LVTLARGLSPAFRLFGGKRTDFLOF 84
Db 80 LRTLARGLSPAYLRFGGTKTDFLIF 104

RESULT 7
Q9MYO0 PRELIMINARY; PRT; 545 AA.
AC Q9MYO0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HEPARANASE.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Kizaki K., Nakano H., Takahashi T., Imai K., Hashizume K.;
RT "Expression of Heparanase mRNA in Bovine Placenta During Gestation.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF281160; AAF87301.2; -
SQ SEQUENCE 545 AA; 61076 MW; FAC4BDFD855B933 CRC64;

Query Match 31.4%; Score 140.5; DB 6; Length 545;
Best Local Similarity 43.0%; Pred. No. 9.5e-08;
Matches 34; Conservative 12; Mismatches 24; Indels 9; Gaps 2;

QY 7 PVDRAAGLKEKTLILLDVSTKNPVTNENFLSLQDPSIHD-GWLDPLSKRLVTLAR 65
Db 36 PADDAAE-----LEFFTERPLHLVSPAFSLFTIDANLATDPRFFTLGSSKRLTLAR 87
QY 66 GLSPAFRLFGGKRTDFLOF 84
Db 88 GLAPAYLRFGGKNGDPLIF 106

RESULT 8
Q90YK5 PRELIMINARY; PRT; 523 AA.
ID Q90YK5
AC Q90YK5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HEPARANASE.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=11387326;
RA Goldsmith O., Zcharia E., Aingorn H., Guatta-Rangini Z., Atzmon R.,
RA Michal I., Pecker I., Mitrani E., Vlodavsky I.;
RT "Expression Pattern and Secretion of Human and Chicken Heparanase Are
RT Determined by Their Signal Peptide Sequence.";
RL J. Biol. Chem. 276:29178-29187(2001).
DR EMBL; AY037007; AAK82648.1; -
SQ SEQUENCE 523 AA; 58386 MW; 8EB0B7B18C9BF881 CRC64;

Query Match 29.6%; Score 132.5; DB 13; Length 523;
Best Local Similarity 48.4%; Pred. No. 7.6e-07;
Matches 31; Conservative 8; Mismatches 24; Indels 1; Gaps 1;

QY 22 LDVSTKNPVTNENFLSLQDPSIHD-GWLDPLSKRLVTLARGLSPAFRLFGGKRTD 80
Db 21 LQLGLREFIGAVSPAFSLITLDASLARDPRFVALLRHPKHLTLASGLSPGFLRFGGTSTD 80
QY 81 FLQF 84
Db 81 FLIF 84

RESULT 9
Q9C5M5 PRELIMINARY; PRT; 304 AA.
ID Q9C5M5
AC Q9C5M5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 33.0 KDA PROTEIN (AT1G27000/T7N9_6).
GN T7N9.6.

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 ON NCBI_taxid=3702;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RT "Full length cDNA of gene T7N9.6 (GI:8778852).";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
 RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Banno F., Dale J.M., Goldsmith A.D., Lee J.M.,
 RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Yamamura Y., Yu G.,
 RA Yu S., Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C.,
 RA Koesema E., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M.,
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
 RA Southwick A., Tracy S.E., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RT "Full length cDNA of gene T7N9.6 (GI:8778852).";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF360143; AA25853.1; -
 DR EMBL; AY057525; AAL09765.1; -
 DR EMBL; AY056367; AAL07253.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 304 AA; 33013 MW; 463DB4579AF52195 CRC64;

Query Match 16.28; Score 72.5; DB 10; Length 304;
 Best Local Similarity 31.08; Pred. No. 3.2; Indels 9; Gaps 3;
 Matches 27; Conservative 15; Mismatches 36; Indels 9; Gaps 3;
 QY 10 RAAGLKEKTLILLDVSTK--NPVRTVNENFLSLQDPSIIH-----DGWLDFLSSKRLV 61
 Db 157 RIONLDDKVEKIDLSKINSQVISARENISLMDLSLNLITGLDGKLDLTLEYKQDV 216
 QY 62 TIARGLSPAFLRFGGKRTDFLQFNLR 88
 Db 217 TNVFMNL-LYNVFGKSTKLPMEQLQ 242

RESULT 10
 Q9LFY5 ID Q9LFY5 PRELIMINARY; PRT; 334 AA.
 AC Q9LFY5;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE T7N9.6.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 ON NCBI_taxid=3702;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,
 RA Shinn P., Altafi H., Bei B., Chin C., Chiou J., Choi E., Conn L.,
 RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
 RA Ecker J.R.;
 RT "Genomic sequence for Arabidopsis thaliana BAC T7N9 from chromosome
 I.";
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
 RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
 RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
 RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
 RA Theologis A., Ecker J.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
 RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
 RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
 RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
 RA Theologis A., Ecker J.;
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC000348; AAF79851.1; -
 SQ SEQUENCE 334 AA; 36323 MW; 82391C5498EF3760 CRC64;

Query Match 16.28; Score 72.5; DB 10; Length 334;
 Best Local Similarity 31.08; Pred. No. 3.5; Indels 9; Gaps 3;
 Matches 27; Conservative 15; Mismatches 36; Indels 9; Gaps 3;
 QY 10 RAAGLKEKTLILLDVSTK--NPVRTVNENFLSLQDPSIIH-----DGWLDFLSSKRLV 61
 Db 163 RIONLDDKVEKIDLSKINSQVISARENISLMDLSLNLITGLDGKLDLTLEYKQDV 222
 QY 62 TIARGLSPAFLRFGGKRTDFLQFNLR 88
 Db 223 TNVFMNL-LYNVFGKSTKLPMEQLQ 248

RESULT 11
 Q9LRC8 ID Q9LRC8 PRELIMINARY; PRT; 527 AA.
 AC Q9LRC8;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE BETA-GLUCURONIDASE.
 GN SGUS.
 OS Scutellaria baicalensis.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

```
OC Asteridae; euasterids I; Lamiales; Lamiaceae; Scutellaria.
OX NCBI_TaxID=65409;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20418130; PubMed=10858442;
RA Sasaki K., Taura F., Shoyama Y., Morimoto S.;
RT "Molecular Characterization of a Novel beta-Glucuronidase from
RT Scutellaria baicalensis Georgi.";
RL J. Biol. Chem. 275:27466-27472(2000).
DR EMBL; AB040072; BAA97804.1; -
DR InterPro; IPR001179; FKBP_PPIase.
DR PROSITE; PS00453; FKBP_PPIASE.1; UNKNOWN.1.
SQ SEQUENCE 527 AA; 58772 MW; ASDE7C423FAIE2B CRC64;

Query Match 15.3%; Score 70.5; DB 10; Length 527;
Best Local Similarity 28.7%; Pred. No. 10;
Matches 25; Conservative 17; Mismatches 28; Indels 17; Gaps 5;

QY 15 KEKTLILLDVSTKNPVTNENFLSLQD---PSIIHDG---W-----LDFSSKRLV 61
Db 26 EETIVKIE---ENPVAQTIDENYVCATLDLWPTKCYGNCPCWCKSSFLNLD-LNNNIIR 81
QY 62 TLARGLSPAFLRFGGKRTDFLOFNLR 88
Db 82 NAVKEFAPLKLRFGGTLDRLVYQTSR 108

RESULT 12
Q9FG13 ID Q9FG13 PRELIMINARY; PRT; 329 AA.
AC Q9FG13;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GB|AAD04946.2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Katoh T., Asanizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002543; BAB11406.1; -
DR InterPro; IPR000379; Est_lip_thioest_actsite.
SQ SEQUENCE 329 AA; 36597 MW; 000140D8BB91126A CRC64;

Query Match 15.3%; Score 68.5; DB 10; Length 329;
Best Local Similarity 25.4%; Pred. No. 10;
Matches 15; Conservative 14; Mismatches 25; Indels 5; Gaps 1;

QY 28 NPVTNENFLSLQDPSIIHDGWLDFLSSKRLVTLARGLSPAFLRFGGKRTDFLOFN 86
Db 246 NPFGPTPTLESISLEPLVIVGSGELLRDR-----AKEYAYKLKMGKRVDTIEFN 299

RESULT 13
Q98PE1 ID Q98PE1 PRELIMINARY; PRT; 694 AA.
AC Q98PE1;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DE 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE EXODEOXYRIBONUCLEASE V ALPHA CHAIN (EC 3.1.11.5).
GN MYPU.7820.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
```

```
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2107;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=UAB CTIP;
RX MEDLINE=21267165; PubMed=11353084;
RA Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA Blanchard A.;
RT "The complete genome sequence of the murine respiratory pathogen
RT Mycoplasma pulmonis.";
RL Nucleic Acids Res. 29:2145-2153(2001).
DR EMBL; AL445565; CACI3955.1; -
DR MypuList; MYPU_7820; -
DR InterPro; IPR003975; Shal_channel.
DR PRINTS; PR01497; SHALCHANNEL.
KW Hydrolase; Complete proteome.
KW SEQUENCE 694 AA; 80774 MW; 61D38974D4465132 CRC64;

Query Match 15.3%; Score 68.5; DB 16; Length 694;
Best Local Similarity 35.4%; Pred. No. 24;
Matches 17; Conservative 8; Mismatches 16; Indels 7; Gaps 1;

QY 1 GDRRLPVDRAAGLKEKTLILLDVST-----KNPVTNENFLSLQ 41
Db 428 GDRQLPAIGAGNLLDLELQNISVNLRSQIFRNKTKDISENFLVQ 475

RESULT 14
Q91MM5 ID Q91MM5 PRELIMINARY; PRT; 305 AA.
AC Q91MM5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE LSDV139 PUTATIVE SER/THR PROTEIN KINASE.
GN LSDV139.
OS lumpy skin disease virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Capripoxvirus.
OX NCBI_TaxID=59509;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEETHLING 2490;
RX MEDLINE=21329495; PubMed=11435593;
RA Tulman E.R., Afonso C.L., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
RT "Genome of lumpy skin disease virus.";
RL J. Virol. 75:7122-7130(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NEETHLING 2490;
RA Tulman E.R., Afonso C.L., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF325528; AAK85100.1; -
KW Kinase.
KW SEQUENCE 305 AA; 35635 MW; F77A0E94273AB816 CRC64;

Query Match 14.7%; Score 66; DB 12; Length 305;
Best Local Similarity 23.0%; Pred. No. 18;
Matches 14; Conservative 17; Mismatches 16; Indels 14; Gaps 2;

QY 20 ILLDVSTKNPVR-----TVNENFLSLQDPSIIHDGWLDFLSSKRLVTLARGLSP 69
Db 163 IAIIDNNKNIYLLDYGLSYRPMINGNHVEYKRDPKKMHNGTIEYTS-----IDMHKGVS 218
QY 70 A 70
Db 219 S 219

RESULT 15
```

017772
ID O17772 PRELIMINARY; PRT; 1391 AA.
AC O17772;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE F09C3.1 PROTEIN.
GN F09C3.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_taxID=6239;
RN [1]
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; Pubmed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL: Z92781; CAB07179.2; ..
DR InterPro; IPR001715; Calponin_hom.
DR InterPro; IPR003247; CH_type.
DR InterPro; IPR000048; IQ.
DR InterPro; IPR001936; RasGAP.
DR InterPro; IPR000593; RasGAP_C.
DR Pfam: PF00612; IQ; 2.
DR ProDom; PD001527; CH_type; 1.
DR ProDom; PD008735; RasGAP_C; 1.
DR SMART; SM00033; CH; 1.
DR SMART; SM00015; IQ; 1.
DR PROSITE; PS50021; CH; 1.
DR PROSITE; PS50018; RAS_GTPASE_ACTIV_2; 1.
SQ SEQUENCE 1391 AA; 159210 MW; 794DF9EBFA2E65AC CRC64;

Query Match 14.6%; Score 65.5; DB 5; Length 1391;
Best Local Similarity 23.5%; Pred. NO. 1.2e+02;
Matches 27; Conservative 12; Mismatches 41; Indels 35; Gaps 3;

QY 8 VDRAGLKEKTLILLDVSTKNP-VRTVNFNLSLQID----- 43
Db 485 VEEAFKLAQVVRINKAKGSEPKVKKELEKLSFKLDGYRELLHWYAKRISEELDASKH 544

QY 44 PSIIHDGWLDF-----LSSKRLVTIARGLSPAPLREGGKRTDFLQFQNL 87
Db 545 PESSQDFWLDHKFPLGTIYVNTSTSKHSTTSQELHPNPLRWNPQLQDFVAYENY 599

Search completed: July 30, 2002, 08:30:43
Job time: 995 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:15:01 ; Search time 53.59 Seconds
(without alignments)
40.109 Million cell updates/sec

Title: US-09-836-461-2_COPY_42_129
Perfect score: 448
Sequence: 1 GDRRPLPVDRAAGLKEKTLI.....PAFLRFGGKRTDFLOFONLR 88

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

```

Database :
Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:**
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:**
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:**
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:**
5: /cgn2_6/ptodata/2/iaa/PCtUS_COMB.pep:**
6: /cgn2_6/ptodata/2/iaa/backfilesl.pep:**

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			Length	DB	ID	Description
	Score	Match					
1	150.5	33.6	532	4	US-09-181-336-15	Sequence 15, Appl	
2	150.5	33.6	543	2	US-08-922-1708-10	Sequence 10, Appl	
3	150.5	33.6	543	2	US-09-071-739B-2	Sequence 2, Appl	
4	150.5	33.6	543	4	US-09-181-336-13	Sequence 13, Appl	
5	150.5	33.6	543	4	US-09-260-038B-2	Sequence 2, Appl	
6	62.5	14.0	732	1	US-08-481-626-2	Sequence 2, Appl	
7	62.5	14.0	732	4	US-08-989-299-4	Sequence 4, Appl	
8	62.5	14.0	1306	4	US-08-989-299-7	Sequence 7, Appl	
9	61	13.6	352	2	US-08-751-474-2	Sequence 2, Appl	
10	61	13.6	649	4	US-09-188-930-305	Sequence 305, App	
11	59	13.2	251	3	US-08-483-807-9	Sequence 9, Appl	
12	57.5	12.8	475	2	US-08-698-407-4	Sequence 4, Appl	
13	57.5	12.8	475	3	US-09-195-855-4	Sequence 4, Appl	
14	57	12.7	211	6	5194596-13	Patent No. 5194596	
15	57	12.7	211	6	5219739-13	Patent No. 5219739	
16	57	12.7	632	2	US-09-354-129-8	Sequence 8, Appl	
17	57	12.7	989	2	US-08-070-301-16	Sequence 16, Appl	
18	56.5	12.6	122	1	US-07-956-700B-37	Sequence 37, Appl	
19	56.5	12.6	122	1	US-08-476-537-37	Sequence 37, Appl	
20	56.5	12.6	122	1	US-08-483-607-37	Sequence 37, Appl	
21	56.5	12.6	122	2	US-08-475-879-37	Sequence 37, Appl	
22	56.5	12.6	2237	1	US-08-354-973-1	Sequence 1, Appl	
23	56	12.5	95	3	US-08-946-329A-78	Sequence 78, Appl	
24	56	12.5	162	4	US-08-858-207A-13	Sequence 263, App	
25	56	12.5	317	1	US-08-221-750A-63	Sequence 13, Appl	
26	55.5	12.4	588	2	US-08-459-346-12	Sequence 12, Appl	
27	55.5	12.4	588	3	US-08-889-419-12	Sequence 12, Appl	

28	55.5	12.4	588	5	PCT-US93-07189-12	Sequence 12, Appl
29	55.5	12.4	911	2	US-08-484-438-10	Sequence 10, Appl
30	55.5	12.4	935	1	US-07-707-367-2	Sequence 2, Appl
31	55.5	12.4	1058	2	US-08-484-438-4	Sequence 4, Appl
32	55.5	12.4	1308	2	US-08-484-438-2	Sequence 2, Appl
33	55	12.3	203	3	US-09-064-703-7	Sequence 7, Appl
34	55	12.3	368	3	US-08-972-902-2	Sequence 2, Appl
35	55	12.3	733	4	US-09-192-983-6	Sequence 6, Appl
36	55	12.3	2482	1	US-08-328-254-6	Sequence 6, Appl
37	55	12.3	3248	1	US-08-353-700-1	Sequence 1, Appl
38	55	12.3	3248	5	PCT-US95-16216-1	Sequence 1, Appl
39	54.5	12.2	207	1	US-08-028-463-3	Sequence 3, Appl
40	54.5	12.2	207	1	US-08-461-836-3	Sequence 3, Appl
41	54	12.1	397	1	US-07-603-133B-9	Sequence 9, Appl
42	54	12.1	448	4	US-08-952-127-21	Sequence 21, Appl
43	54	12.1	3056	1	US-08-508-836A-8	Sequence 8, Appl
44	54	12.1	3056	2	US-08-629-001A-3	Sequence 3, Appl
45	54	12.1	3056	2	US-08-874-266-2	Sequence 2, Appl

ALIGNMENTS

```

RESULT      1
US-09-181-336-15
; Sequence 15, Application US/09181336A
; Patent No. 6242238
; GENERAL INFORMATION:
; APPLICANT: FREEMAN, Craig Geoffrey
; APPLICANT: PARISH, Richard
; APPLICANT: HAMDORF, Brenton James
; APPLICANT: HULETT, Mark Darren
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN
; TITLE OF INVENTION: ENDOGLUCURONIDASE AND USES THEREFOR
; FILE REFERENCE: 032503-004
; CURRENT APPLICATION NUMBER: US/09/181.336A
; CURRENT FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: AU PP0062
; EARLIER FILING DATE: 1997-10-28
; EARLIER APPLICATION NUMBER: AU PP0812
; EARLIER FILING DATE: 1997-12-09
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 532
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-181-336-15

```

```
Query Match          33.6%; Score 150.5; DB 4; Length 532;
Best Local Similarity 43.5%; Pred. No. 1.2e-11;
Matches 37; Conservative 17; Mismatches 28; Indels 3; Gaps 2;
```

QY	1	GRRPLPVDRAAGKCKTLLLDVSTKNPVRTNVNENFLSLQLDPSIIHD-GWLDPLLSKR	59
		: : : : : : : : : : :	:
Dd	22	GPLSPGALPRPA--QAQDVVDLDFEOTPEPLHLVSPSFLSVTDANLATDPREFILLGS	79
QY	60	LVTIARGLSPAFLRFEGCKRDTDFLOF	84
			:
Dd	80	LRTIARGLSPAYLRFEGTKTKTDFLI	104

```

RESULT      2
US-08-922-170B-10
; Sequence 10, Application us/08922170B
; Patent No. 5968822
; GENERAL INFORMATION:
; APPLICANT: Iris Pecker, Israel Vlodavsky and Elena
; APPLICANT: Feinstein
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A POLYPEPTIDE
; TITLE OF INVENTION: HAVING HEPARANASE ACTIVITY AND EXPRESSION OF
; TITLE OF INVENTION: SAME IN TRANSFUSED CELLS
;

```


Db 22 GPLSPGALPRPA--QAQDVVDLDFFTQEPHLVSPSFLSVTIDANLATDPRFLLILGSPK 79
QY 60 LVTLAGLSPAFLRFGKRTDQLQF 84
Db 80 LRTLARGLSPAYLRFGGTKTDLFIF 104

RESULT 5
US-09-260-038B-2
; Sequence 2, Application US/09260038B
; Patent No. 6348344
; GENERAL INFORMATION:
; APPLICANT: Maty Ayal-Hershkovitz et al.
; TITLE OF INVENTION: GENETICALLY MODIFIED CELLS AND METHODS FOR
; EXPRESSING RECOMBINANT HEPARANASE
; AND METHODS OF PURIFYING SAME
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
; STREET: 2001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; SOFTWARE: Word for Windows version 2.0 converted to
; an ASCII file
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/260.038B
; FILING DATE: 02-Mar-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/071,618
; FILING DATE: May 1, 1998
; APPLICATION NUMBER: 09/071,739
; FILING DATE: May 1, 1998
; APPLICATION NUMBER: 08/922,180
; FILING DATE: September 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedman, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 910/16
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 972-3-562553
; TELEFAX: 972-3-562554
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-260-038B-2

Query Match 33.6%; Score 150.5; DB 4; Length 543;
Best Local Similarity 43.5%; Pred. No. 1.3e-11;
Matches 37; Conservative 17; Mismatches 28; Indels 3; Gaps 2;
QY 1 GDRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQDPSIIHD-CWMLDFLSKR 59
Db 22 GPLSPGALPRPA--QAQDVVDLDFFTQEPHLVSPSFLSVTIDANLATDPRFLLILGSPK 79
QY 60 LVTLAGLSPAFLRFGKRTDQLQF 84
Db 80 LRTLARGLSPAYLRFGGTKTDLFIF 104

RESULT 6
US-08-481-626-2
; Sequence 2, Application US/08481626
; Patent No. 5801040
; GENERAL INFORMATION:
; APPLICANT: Soubrier, Florent
; APPLICANT: Albenc-Gelas, Francois
; APPLICANT: Hubert, Christine
; APPLICANT: Corvol, Pierre
; TITLE OF INVENTION: Nucleic Acid Coding for the Human
; Tissue Angiotensin Converting Enzyme (ACE) and its
; Uses, Especially for the In Vitro Screening for this
; Enzyme in the Organism
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481,626
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/656,183
; FILING DATE: 04-Mar-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 89-09062
; FILING DATE: 05-JUL-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 04958-0006-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 732 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-481-626-2

Query Match 14.0%; Score 62.5; DB 1; Length 732;
Best Local Similarity 32.4%; Pred. No. 8.8;
Matches 22; Conservative 8; Mismatches 19; Indels 3; Gaps 3;
QY 2 DRPLPVDRAAGLKEKTLILLDVSTKNPVRTV-NENFLSLQDPSIIH----- 48
Db 154 ERAALP---AQELEBYNKILLDMETTVSVATVCHPNGSCQLQLEPDLTNVMTSRKYEDLL 210
QY 49 ---DGWLD 53
Db 211 WAWEGWRD 218
RESULT 7
US-08-989-299-4
; Sequence 4, Application US/08989299
; Patent No. 6194556
; GENERAL INFORMATION:

APPLICANT: Acton, Susan L.
APPLICANT: Robinson, Keith E.
TITLE OF INVENTION: ANGIOTENSIN CONVERTING ENZYME HOMOLOG
TITLE OF INVENTION: AND THERAPEUTIC AND DIAGNOSTIC USES THEREFOR
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 11-DEC-1997
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Arnold E., Beth
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: MIA-025.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1306 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-989-299-7

Query Match 14.0%; Score 62.5; DB 4; Length 1306;
Best Local Similarity 32.4%; Pred. No. 19;
Matches 22; Conservative 8; Mismatches 19; Indels 19; Gaps 3;

QY 2 DRRLPVDRAAGLKEKTLILLDVSTKNPVRTV-NENFLSLQDPDSIIH----- 48
Db 154 ERAALP---AQEEYVKILLDMETTSVATVCHPNCGSCQLQEPDLTNVMTSRKYEDLL 210
QY 49 ----DGWLD 53
Db 211 WAVEGWDRD 218

RESULT 8
US-08-989-299-7
; Sequence 7, Application US/08989299
; Patent No. 6194556
; GENERAL INFORMATION:
; APPLICANT: Acton, Susan L.
; APPLICANT: Robinson, Keith E.
; TITLE OF INVENTION: ANGIOTENSIN CONVERTING ENZYME HOMOLOG
; TITLE OF INVENTION: AND THERAPEUTIC AND DIAGNOSTIC USES THEREFOR
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

APPLICANT: Acton, Susan L.
APPLICANT: Robinson, Keith E.
TITLE OF INVENTION: ANGIOTENSIN CONVERTING ENZYME HOMOLOG
TITLE OF INVENTION: AND THERAPEUTIC AND DIAGNOSTIC USES THEREFOR
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 11-DEC-1997
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Arnold E., Beth
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: MIA-025.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 732 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-989-299-4

Query Match 14.0%; Score 62.5; DB 4; Length 732;
Best Local Similarity 32.4%; Pred. No. 8.8; 19; Indels 19; Gaps 3;
Matches 22; Conservative 8; Mismatches 19; Indels 19; Gaps 3;

QY 2 DRRLPVDRAAGLKEKTLILLDVSTKNPVRTV-NENFLSLQDPDSIIH----- 48
Db 154 ERAALP---AQEEYVKILLDMETTSVATVCHPNCGSCQLQEPDLTNVMTSRKYEDLL 210
QY 49 ----DGWLD 53
Db 211 WAVEGWDRD 218

RESULT 9
US-08-751-474-2
; Sequence 2, Application US/08751474
; Patent No. 5821335
; GENERAL INFORMATION:
; APPLICANT: Hoskins, JoAnn
; APPLICANT: Skatrud, Paul L.
; TITLE OF INVENTION: Biosynthetic Gene murG From
; TITLE OF INVENTION: Streptococcus Pneumoniae
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-9902
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 352 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-751-474-2


```
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,857
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-504
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 251 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-483-857-9

Query Match 13.2%; Score 59; DB 3; Length 251;
Best Local Similarity 36.0%; Pred. No. 6;
Matches 18; Conservative 7; Mismatches 21; Indels 4; Gaps 2;

QY 4 RPLPVDRAAGLKERTLLLDVSTKNPVRTVNE-NFLSLQLDPSIIHDGWL 52
Db 42 RAVEPD---GLSLATRTLVASTPGADWTPCOLDYTSRPLDVALQDQDGL 88

RESULT 12
US-08-698-407-4
; Sequence 4, Application US/08698407
; Patent No. 5856128
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN NUCLEIC ACID BINDING PROTEIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/698,407
; FILING DATE: Filed Herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0116 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 475 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 267188

Query Match 13.6%; Score 61; DB 2; Length 352;
Best Local Similarity 32.3%; Pred. No. 5.2;
Matches 20; Conservative 8; Mismatches 28; Indels 6; Gaps 3;

QY 32 TVNENFLSLQLDPSIIHDGW-LDFLSKRLV----TLARGLSPAFLEFG-GKRTDFLQFQ 85
Db 11 TVGHVTLNLLMLPKFIEDGWEVHYIGDKRGIEHQEILKSLGDLVTFHSHATGKLRREFSQ 70

QY 86 NL 87
Db 71 NM 72

RESULT 10
US-09-188-930-305
; Sequence 305, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 305
; LENGTH: 649
; TYPE: PRT
; ORGANISM: Mouse
; US-09-188-930-305

Query Match 13.6%; Score 61; DB 4; Length 649;
Best Local Similarity 30.0%; Pred. No. 12;
Matches 27; Conservative 11; Mismatches 34; Indels 18; Gaps 4;

QY 5 PLPVDRAAGLKERTLLLDVSTKNPVR-----TVNENFLSLQLDPSIIHDGWLDFL--SS 57
Db 382 PAPVTQPDIKNEKLKIDQRTTGSPSRKTLITVK-----SVTPDTHISWRLLPMTA 435

QY 58 KRLVTLARGLSPAF-----LRFGGKRTDFL 82
Db 436 LRLSLKLGHSAPAFSGSITETIVTGERSEYL 465

RESULT 11
US-08-483-857-9
; Sequence 9, Application US/08483857
; Patent No. 6020125
; GENERAL INFORMATION:
; APPLICANT: Chan, Voon Loong
; APPLICANT: Louie, Helena
; TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
; TITLE OF INVENTION: CAMPYLOBACTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

US-08-698-407-4

Query Match 12.8%; Score 57.5; DB 2; Length 475;
Best Local Similarity 21.8%; Pred. No. 22;
Matches 19; Conservative 15; Mismatches 24; Indels 29; Gaps 2;
QY 1 GDRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNEFLSLQLODPSIIHDGWLDFLSSKRL 60
Db 326 GDKLLVQASVGAKNATLV-----SPPSTINQTPVTLOVP----- 361

QY 61 VTLARGLSPAFLRFGGKRTDFLQFQNL 87
Db 362 -----GLMSSQVQMGHPTFVLC LMNM 383

RESULT 13
US-09-195-855-4
; Sequence 4, Application US/09195855
; Patent No. 6015788
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN NUCLEIC ACID BINDING PROTEIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA: US/09/195,855
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/698,407
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0116 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 475 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 267188
US-09-195-855-4

Query Match 12.8%; Score 57.5; DB 3; Length 475;
Best Local Similarity 21.8%; Pred. No. 22;
Matches 19; Conservative 15; Mismatches 24; Indels 29; Gaps 2;
QY 1 GDRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNEFLSLQLODPSIIHDGWLDFLSSKRL 60
Db 326 GDKLLVQASVGAKNATLV-----SPPSTINQTPVTLOVP----- 361

QY 61 VTLARGLSPAFLRFGGKRTDFLQFQNL 87
Db 362 -----GLMSSQVQMGHPTFVLC LMNM 383
RESULT 14
5194596-13
; Patent No. 5194596
; APPLICANT: TISCHER, EDMUND G.; ABRAHAM, JUDITH A.; FIDDES, JOHN
; C.; MITCHELL, RICHARD L.
; TITLE OF INVENTION: PRODUCTION OF VASCULAR ENDOTHELIAL CELL
; GROWTH FACTOR
; NUMBER OF SEQUENCES: 32
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/450,883
; FILING DATE: 14-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 387,545
; FILING DATE: 27-JUL-1989
; SEQ ID NO:13:
; LENGTH: 211
5194596-13

Query Match 12.7%; Score 57; DB 6; Length 211;
Best Local Similarity 23.9%; Pred. No. 8.8;
Matches 17; Conservative 13; Mismatches 17; Indels 24; Gaps 2;
QY 2 DRRLPVDRAAGLKE-----KTLILLDVSTKNPVRTVNEFL----- 38
Db 76 EKRLPIRRKRRIEAEVPAVCKTRIVIEIPRSQVDPTSANFLIWPVPCVEVKRCTGCNT 135

QY 39 -SLQLDPSIIH 48
Db 136 SSVKQCPSRVH 146

RESULT 15
5219739-13
; Patent No. 5219739
; APPLICANT: TISCHER, EDMUND G.; ABRAHAM, JUDITH A.; FIDDES,
; JOHN C.; MITCHELL, RICHARD L.
; TITLE OF INVENTION: DNA SEQUENCES ENCODING BVGEF120 AND
; BVGEF 121 AND METHODS FOR THE PRODUCTION OF BOVINE AND HUMAN
; VASCULAR ENDOTHELIAL CELL GROWTH FACTORS, BVGEF120 AND BVGEF121
; NUMBER OF SEQUENCES: 40
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/559,041
; FILING DATE: 27-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 450,883
; FILING DATE: 14-DEC-1989
; APPLICATION NUMBER: 387,545
; FILING DATE: 27-JUL-1989
; SEQ ID NO:13:
; LENGTH: 211
5219739-13

Query Match 12.7%; Score 57; DB 6; Length 211;
Best Local Similarity 23.9%; Pred. No. 8.8;
Matches 17; Conservative 13; Mismatches 17; Indels 24; Gaps 2;
QY 2 DRRLPVDRAAGLKE-----KTLILLDVSTKNPVRTVNEFL----- 38
Db 76 EKRLPIRRKRRIEAEVPAVCKTRIVIEIPRSQVDPTSANFLIWPVPCVEVKRCTGCNT 135

QY 39 -SLQLDPSIIH 48
Db 136 SSVKQCPSRVH 146

Search completed: July 30, 2002, 08:15:02
Job time: 284 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:14:02 ; Search time 141.47 Seconds
(without alignments)
94.217 Million cell updates/sec

Title: US-09-836-461-2_COPY_42_161
Perfect score: 621
Sequence: 1 GDRPLPVDRAAGLKEKTLI.....YLLKNYEDDIVRSDVALDKQ 120

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	621	100.0	262	22	AA24147 Human EST encoded
2	621	100.0	534	22	AA885216 Heparanase-like pr
3	621	100.0	592	22	AAU07424 Human heparanase-1
4	621	100.0	592	22	AA885215 Heparanase-like pr
5	621	100.0	592	22	AA881062 Human Heparanase-2
6	621	100.0	592	22	AA97632 Human heparanase,
7	615	99.0	492	22	AA884664 Amino acid sequenc
8	566	91.1	439	22	AAU07423 Human heparanase-1
9	566	91.1	480	22	AAU07418 Novel human extrac
10	566	91.1	480	22	AA885217 Heparanase-like pr
11	566	91.1	480	22	AA97634 Human heparanase,

12	565.5	91.1	538	22	AA97633 Human heparanase,
13	181.5	29.2	535	21	AA808851 A murine heparanas
14	156.5	25.2	532	20	AAV17083 Seq ID No: 15 of W
15	156.5	25.2	543	20	AAV17082 Human heparanase e
16	156.5	25.2	543	20	AAV02345 A human heparanase
17	156.5	25.2	543	21	AA808849 Amino acid sequenc
18	156.5	25.2	543	21	AAV57590 Human heparanase. p
19	156.5	25.2	543	21	AAV52990 Human heparanase. p
20	156.5	25.2	543	22	AA886206 Human heparanase i
21	156.5	25.2	543	22	AA888361 Human membrane or
22	156.5	25.2	543	22	AA97635 Human heparanase p
23	156.5	25.2	588	20	AA930124 A human protein wi
24	156.5	25.2	592	20	AAV02346 A human heparanase
25	156.5	25.2	592	21	AA808850 Amino acid sequenc
26	155.5	25.0	530	20	AAV34173 Human pre-prohepar
27	93	15.0	32	20	AAV34175 Human pre-prohepar
28	86.5	13.9	488	22	AA831469 Amino acid sequenc
29	84.5	13.6	280	21	AA831469 Arabidopsis thalia
30	84.5	13.6	302	21	AA831469 Arabidopsis thalia
31	84.5	13.6	304	21	AA831469 Arabidopsis thalia
32	82.5	13.3	280	21	AA831469 Arabidopsis thalia
33	82.5	13.3	302	21	AA831469 Arabidopsis thalia
34	82.5	13.3	304	21	AA831469 Arabidopsis thalia
35	82.5	13.3	488	22	AA831472 Amino acid sequenc
36	80.5	13.0	488	22	AA831470 Amino acid sequenc
37	76.5	12.3	488	22	AA831471 Amino acid sequenc
38	75	12.1	241	22	AB869355 Drosophila melanog
39	74	11.9	402	22	ABG09311 Novel human diago
40	73.5	11.8	445	22	AA840763 Human polypeptide
41	73.5	11.8	445	22	AA840763 Human polypeptide
42	73.5	11.8	445	22	AA840765 Human polypeptide
43	73.5	11.8	641	22	AA838978 Human polypeptide
44	73.5	11.8	698	22	AA838977 Human polypeptide
45	73.5	11.8	698	22	AA839032 Human polypeptide

ALIGNMENTS

RESULT	1
AA24147	
ID	AA24147 standard; Protein; 262 AA.
XX	AA24147;
AC	AC
XX	XX
DT	12-OCT-2001 (first entry)
XX	XX
DE	Human EST encoded protein SEQ ID NO: 1672.
XX	XX
KW	Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW	tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW	diagnostics; forensic test; gene mapping; genetic disorder;
KW	biodiversity; gene therapy; nutrition.
XX	XX
OS	Homo sapiens.
XX	XX
PN	WO200154477-A2.
XX	XX
PD	02-AUG-2001.
XX	XX
PF	25-JAN-2001; 2001WO-US02687.
XX	XX
PR	25-JAN-2000; 2000US-0491404.
PR	17-JUL-2000; 2000US-0617746.
PR	03-AUG-2000; 2000US-0631451.
PR	15-SEP-2000; 2000US-0663870.
XX	XX
PA	(HYSE-) HYSEQ INC.
XX	XX
PI	Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI	Cao Y, Drmanac RA, Zhang J, Werhman T;
XX	XX
DR	WPI; 2001-476164/51.

DR N-PSDB; AAH98806.
 XX Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use -
 XX
 XX
 PS Claim 20; Page 1122-1123; 1275pp; English.
 XX
 CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a
 CC protein of the invention.
 XX
 XX Sequence 262 AA;
 SQ

Query Match 100.0%; Score 621; DB 22; Length 262;
 Best Local Similarity 100.0%; Pred. No. 9.1e-65;
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDRRLPVDRAAGLKEKTLILLDVTSTKNPVRTVNENFLSLQDPSIIHDGWLDFLSSKRL 60
 DB 42 GDRRLPVDRAAGLKEKTLILLDVTSTKNPVRTVNENFLSLQDPSIIHDGWLDFLSSKRL 101
 QY 61 VTARGLSPAFLRFGGKRTDFLQFONLRNPAKSRGGPGPDYVLYKNYEDDIVRSDVALDKQ 120
 DB 102 VTARGLSPAFLRFGGKRTDFLQFONLRNPAKSRGGPGPDYVLYKNYEDDIVRSDVALDKQ 161

RESULT 2
 AAB85216
 ID AAB85216 standard; Protein; 534 AA.
 XX
 AC AAB85216;
 XX
 DT 07-SEP-2001 (first entry)
 XX
 DE Heparanase-like protein Hpa2 splice variant #2.
 XX
 KW Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;
 KW cytosolic; neuroprotective; cerebroprotective; immunosuppressive;
 KW antiproliferative; antidiabetic; antiinflammatory; antiarthritic; antiasthmatic;
 KW antidiabetic; antiarteriosclerotic; vulnerary.
 XX
 OS Homo sapiens.
 XX
 FN WO200146392-A2.
 XX
 PD 28-JUN-2001.
 XX
 PF 21-DEC-2000; 2000WO-GB04963.
 XX
 PR 22-DEC-1999; 99GB-0030392.
 XX
 PR 07-APR-2000; 2000GB-0008713.
 XX
 PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
 XX
 XX McKenzie EA, Stamps AC, Terrett JA, Tyson KL;
 PI WPI: 2001-418056/44.
 XX
 DR N-PSDB; AAH22672.
 XX
 XX Novel homologs of heparanase, present in three splice variants, useful
 PT for identifying agents that modulate heparanase, useful in the
 PT treatment and/or prophylaxis of abnormal levels of heparanase -
 XX
 PS Claim 1; Fig 2; 97pp; English.
 XX
 CC The invention provides a homologue to heparanase which is present in
 CC three splice variants. The heparanase homologue polypeptide is useful in

CC the treatment of a human or non-human animal or for use in diagnosis.
 CC Vectors comprising the heparanase homologue polynucleotides are useful in
 CC the transformation or transfection of a prokaryotic or eukaryotic host.
 CC The modulators of the polypeptide are useful in the manufacture of a
 CC medicament for the treatment and/or prophylaxis of a condition/disease
 CC associated with abnormal levels of the heparanase homologue, including
 CC cancer, central nervous system (CNS) and neurodegenerative diseases,
 CC cardiovascular diseases such as stenosis following angioplasty and
 CC atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus,
 CC allografts, inflammatory diseases, arthritis, vascular restenosis,
 CC tumour growth and progression, asthma, Alzheimer's disease, diabetic
 CC retinopathy, wound healing and inflammation. The polypeptide is also
 CC useful in diagnosis and research. The present sequence represents the
 CC amino acid sequence of the mid-sized splice variant of the heparanase-
 CC like protein Hpa2 of the invention.
 XX
 XX Sequence 534 AA;
 SQ

Query Match 100.0%; Score 621; DB 22; Length 534;
 Best Local Similarity 100.0%; Pred. No. 2.3e-64;
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDRRLPVDRAAGLKEKTLILLDVTSTKNPVRTVNENFLSLQDPSIIHDGWLDFLSSKRL 60
 DB 42 GDRRLPVDRAAGLKEKTLILLDVTSTKNPVRTVNENFLSLQDPSIIHDGWLDFLSSKRL 101
 QY 61 VTARGLSPAFLRFGGKRTDFLQFONLRNPAKSRGGPGPDYVLYKNYEDDIVRSDVALDKQ 120
 DB 102 VTARGLSPAFLRFGGKRTDFLQFONLRNPAKSRGGPGPDYVLYKNYEDDIVRSDVALDKQ 161

RESULT 3
 AAU07424
 ID AAU07424 standard; Protein; 592 AA.
 XX
 AC AAU07424;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Human heparanase-like protein splice variant #1.
 XX
 KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
 KW antiproliferative; cardiant; vasotropic; cerebroprotective; nootropic;
 KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
 KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;
 KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
 KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; sunburn;
 KW nervous system disorder; Alzheimer's disease; ocular disorder; infection;
 KW wound healing; food additive; heparanase.
 XX
 OS Homo sapiens.
 XX
 FN WO200179253-A1.
 XX
 PD 25-OCT-2001.
 XX
 PF 11-APR-2001; 2001WO-US11643.
 XX
 PR 18-APR-2000; 2000US-198123P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Fiscella M, Shi Y, Ebner R, Ruben SM;
 PI WPI: 2001-611720/70.
 XX
 DR N-PSDB; AAS13848.
 XX
 PT New nucleic acids encoding extracellular matrix polypeptides, for
 PT diagnosing, treating, preventing or ameliorating human disorders and
 PT disease, such as, autoimmune, hyperproliferative or cardiovascular
 PT disorders -
 XX

XX PF 11-SEP-2000; 2000WO-EP08837.
 XX XX 23-SEP-1999; 99EP-0118805.
 PR 07-JUL-2000; 2000EP-0114649.
 XX (MERE) MERCK PATENT GMBH.
 XX PA Duecker K, Sirrenberg C;
 XX PI WPI; 2001-308089/32.
 DR N-PSDB; AAF86101.
 XX
 PT New heparanase-2 polypeptide useful in diagnosing (the susceptibility
 of a subject to) and as vaccines against e.g. autoimmune disorders,
 PT cardiovascular disease, cancer, diabetes, ischemia, sepsis, stroke, or
 PT thrombosis -
 XX
 PS Claim 1; Page 42-43; 45pp; English.
 XX
 CC This invention relates to a human heparanase-2 protein and the cDNA
 CC sequence encoding it. Heparanase-2 is a member of the endoglucuronidase
 CC family of polypeptides and it degrades heparan sulphate proteoglycans
 CC HSPGs (ubiquitous macromolecules of cell surfaces, basement membranes and
 CC the extracellular matrix). HSPGs support the vascular endothelium and
 CC stabilise the structure of the capillary wall. Heparanases may be
 CC associated with neovascularization and metastasis related to malignant
 CC tumour formation. Heparanase-2 polynucleotides and proteins are useful as
 CC vaccines for inducing an immunological response against autoimmune
 CC disorders, blood coagulation disorders, cancer, diabetes, ischaemia,
 CC sepsis, stroke, cardiovascular diseases, or thrombosis, as well as in
 CC diagnosing (the susceptibility of a subject to) these diseases.
 CC Heparanase-2 fragments may be used as immunogens to produce antibodies
 CC immunospecific to the polypeptides, and to identify membrane bound
 CC soluble receptors, agonists or antagonists that compete with the binding
 CC of the polypeptide to the receptors. An antibody specific for
 CC heparanase-2 can be used in the diagnosis of the above diseases and in
 CC isolating or identifying clones expressing heparanase-2. The present
 CC sequence represents heparanase-2. Three regions of heparanase-2 with high
 CC immunogenicity (immunogenic epitopes) can be used to raise antibodies
 CC against heparanase-2.
 XX
 SQ Sequence 592 AA;

Query Match 100.0%; Score 621; DB 22; Length 592;
 Best Local Similarity 100.0%; Pred. No. 2.7e-64;
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDRRLPVDRAAGLKEKTLILLDVSSTKPNVTVNENFLSLQDLPSIIHDGWLDFLSSKRL 60
 DB 42 GDRRLPVDRAAGLKEKTLILLDVSSTKPNVTVNENFLSLQDLPSIIHDGWLDFLSSKRL 101
 QY 61 VTLAGLSPAFLRFGGKRTDFLQFNLRNPAKSRGGPGPDYLYKNYEDDIVRSDVALDKQ 120
 DB 102 VTLAGLSPAFLRFGGKRTDFLQFNLRNPAKSRGGPGPDYLYKNYEDDIVRSDVALDKQ 161

RESULT 6
 AAY97632
 ID AAY97632 standard; Protein; 592 AA.
 XX
 AC AAY97632;
 XX
 XX 20-APR-2001 (first entry)
 DT Human heparanase, hnhp1, protein sequence.
 DE
 DE Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;
 KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
 KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
 KW gene therapy; human.
 XX

OS Homo sapiens.
 XX WO200100643-A2.
 XX 04-JAN-2001.
 PD
 XX 19-JUN-2000; 2000WO-IL00358.
 XX 25-JUN-1999; 99US-0140801.
 XX (INST-) INSIGHT STRATEGY & MARKETING LTD.
 PA Pecker I, Michal I, Itzhaki H;
 XX WPI; 2001-137930/14.
 DR N-PSDB; AAA91097.
 XX
 PT New polynucleotides and polypeptides that are distantly homologous to
 PT heparanase, useful in wound healing, as well as in gene therapy
 PT protocols for angiogenesis, restenosis, atherosclerosis, or
 PT inflammation -
 XX
 PS Claim 10; Fig 1; 67pp; English.
 XX
 CC This sequence represents a heparanase of the invention.
 CC The heparanase DNA and protein sequences are useful in wound healing,
 CC angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary
 CC diseases, neurodegenerative diseases (such as Scrape, Alzheimer's
 CC disease, and Creutzfeldt-Jakob disease) or viral infections. The
 CC heparanase coding sequence is particularly useful in gene therapy.
 XX
 SQ Sequence 592 AA;

Query Match 100.0%; Score 621; DB 22; Length 592;
 Best Local Similarity 100.0%; Pred. No. 2.7e-64;
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDRRLPVDRAAGLKEKTLILLDVSSTKPNVTVNENFLSLQDLPSIIHDGWLDFLSSKRL 60
 DB 42 GDRRLPVDRAAGLKEKTLILLDVSSTKPNVTVNENFLSLQDLPSIIHDGWLDFLSSKRL 101
 QY 61 VTLAGLSPAFLRFGGKRTDFLQFNLRNPAKSRGGPGPDYLYKNYEDDIVRSDVALDKQ 120
 DB 102 VTLAGLSPAFLRFGGKRTDFLQFNLRNPAKSRGGPGPDYLYKNYEDDIVRSDVALDKQ 161

RESULT 7
 AAB84664
 ID AAB84664 standard; Protein; 492 AA.
 XX
 AC AAB84664;
 XX
 XX 05-SEP-2001 (first entry)
 DT Amino acid sequence of human heparanase-like polypeptide.
 DE
 DE Human; heparanase-like polypeptide; gene therapy; cancer; angiogenesis;
 KW trauma; autoimmune disease; skin disease; cardiovascular disease;
 KW nervous system disease; inflammation; arthritis; genitalia;
 KW male fertility; erectile dysfunction.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 407 /note= "unspecified residue encoded by KCA"
 FT
 XX WO200148161-A2.
 XX 05-JUL-2001.
 PD
 XX 18-DEC-2000; 2000WO-EP12909.
 PF


```
XX PR 23-DEC-1999; 99EP-0125831.
XX XX (SCHD ) SCHERING AG.
XX PA
XX XX Siemeister G, Weiss B;
XX PI
XX XX WPT; 2001-418259/44.
XX DR N-PSDB; AAH28347.
XX XX
XX XX Human Heparanase-like polynucleotide encoding polypeptides useful for
PT modulating expression of the polypeptide and for treating cancer,
PT cancer metastasis, aberrant angiogenesis by gene therapy technique -
XX XX
XX PS Claim 9; Page 30; 30pp; English.
XX XX
XX CC The present sequence represents a human heparanase-like polypeptide.
XX CC Heparanase-like polynucleotides are useful as a source of probes,
XX CC primers and antisense molecules, and in gene therapy. Heparanase-like
XX CC polynucleotides and polypeptides are useful for treating several
XX CC disorders e.g., cancer, cancer metastasis. The oligonucleotides are
XX CC also useful as diagnostic markers for the diagnosis of disorder such
XX CC as cancer, cancer metastasis and aberrant angiogenesis. They may also
XX CC act as diagnostic markers for diagnosis of disorder such as cancer,
XX CC cancer metastasis and aberrant angiogenesis. The heparanase polypeptides
XX CC and polynucleotides are also useful for treating trauma, autoimmune
XX CC diseases, skin diseases, cardiovascular diseases, nervous system
XX CC diseases, and inflammation including arthritis. Since the polynucleotide
XX CC is preferentially expressed in male genitalia, modulation of its
XX CC expression and/or activity may be used for medical intervention in male
XX CC genitalia function that is male fertility control, erectile dysfunction.
XX XX
XX SQ Sequence 492 AA;

Query Match 99.08; Score 615; DB 22; Length 492;
Best Local Similarity 100.0%; Pred. No. 1.1e-63;
Matches 119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 DRRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNFNLSLQLDPSIIHDGWLDFLSSKRLV 61
DB 1 drrplpvdraaglkectllildvstknpvrtnvnenflslqldpslihdgwlfdlsskrlv 60

QY 62 TLARGLSPAFURFGGKRTDFLQFQNLNRPKSRGPGPDYLYKNVEDDIVRSDVALDKQ 120
DB 61 tlarglspafirfggkrtdfiqfnlnrpaksgpgpdylyknyeddivrsvdvaldkq 119

RESULT 8
AAU07423
ID AAU07423 standard; Protein; 439 AA.
XX AC
XX AC AAU07423;
XX DT 18-DEC-2001 (first entry)
XX XX
XX DE Human heparanase-like protein.
XX XX
XX KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW antiproliferative; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;
KW nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;
XX wound healing; food additive; heparanase.
XX XX
XX OS Homo sapiens.
XX XX
XX PN WO200179253-A1.
XX XX
XX PD 25-OCT-2001.
XX XX
```

```
PF 11-APR-2001; 2001WO-US11643.
XX XX
XX PR 18-APR-2000; 2000US-198123P.
XX XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX XX
XX PI Fiscella M, Shi Y, Ebner R, Ruben SM;
XX XX WPT; 2001-611720/70.
XX DR
XX XX
XX PT New nucleic acids encoding extracellular matrix polypeptides, for
PT diagnosing, treating, preventing or ameliorating human disorders and
PT disease, such as, autoimmune, hyperproliferative or cardiovascular
PT disorders -
XX XX
XX PS Disclosure; Page 13-14; 308pp; English.
XX XX
XX CC The invention relates to novel isolated polynucleotides (I) encoding
XX CC extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by
XX CC (I) are used to prevent, treat or ameliorate a medical condition in e.g.
XX CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They
XX CC are also used in diagnosing a pathological condition or susceptibility to
XX CC a pathological condition. The antibodies to the polypeptides can also be
XX CC used in alleviating symptoms associated with the disorders and in
XX CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
XX CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
XX CC include autoimmune diseases e.g. rheumatoid arthritis,
XX CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
XX CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
XX CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
XX CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
XX CC and ocular disorders e.g. corneal infection. The polypeptides can also
XX CC be used to aid wound healing and epithelial cell proliferation, to
XX CC prevent skin aging due to sunburn, to maintain organs before
XX CC transplantation, for supporting cell culture of primary tissues, to
XX CC regenerate tissues and in chemotaxis. The polypeptides can also be used
XX CC as a food additive or preservative to increase or decrease storage
XX CC capabilities. The present sequence represents the amino acid sequence
XX CC of human heparanase-like protein.
XX XX
XX SQ Sequence 439 AA;

Query Match 91.1%; Score 566; DB 22; Length 439;
Best Local Similarity 99.1%; Pred. No. 5.4e-58;
Matches 108; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNFNLSLQLDPSIIHDGWLDFLSSKRL 60
DB 1 gdrplpvdraaglkectllildvstknpvrtnvnenflslqldpslihdgwlfdlsskrl 60

QY 61 VTLARGLSPAFURFGGKRTDFLQFQNLNRPKSRGPGPDYLYKNVEDD 109
DB 61 vtlarglspafirfggkrtdfiqfnlnrpaksgpgpdylyknyede 109

RESULT 9
AAU07418
ID AAU07418 standard; Protein; 480 AA.
XX XX
XX AC AAU07418;
XX XX
XX DT 18-DEC-2001 (first entry)
XX XX
XX DE Novel human extracellular matrix (ECM) protein #1.
XX XX
XX KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW antiproliferative; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;
KW nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;
```


DE Human heparanase, hnhp1 pn5 form, protein sequence.
XX
KW Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;
KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
KW gene therapy; human.
XX
OS Homo sapiens.
XX
PN WO200100643-A2.
XX
PD 04-JAN-2001.
XX
PF 19-JUN-2000; 2000WO-IL00358.
XX
PR 25-JUN-1999; 99US-0140801.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI Pecker I, Michal I, Itzhaki H;
XX
DR WPI; 2001-137930/14.
XX
DR N-PSDB; AAA91099.
XX
XX New polynucleotides and polypeptides that are distantly homologous to
PT heparanase, useful in wound healing, as well as in gene therapy
PT protocols for angiogenesis, restenosis, atherosclerosis, or
PT inflammation -
XX
PS Claim 10; Page 63; 67pp; English.
XX
CC This sequence represents a heparanase of the invention.
CC The heparanase DNA and protein sequences are useful in wound healing,
CC angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary
CC diseases, neurodegenerative diseases (such as Scrape, Alzheimer's
CC disease, and Creutzfeldt-Jakob disease) or viral infections. The
CC heparanase coding sequence is particularly useful in gene therapy.
XX
SQ Sequence 480 AA;

Query Match 91.1%; Score 566; DB 22; Length 480;
Best Local Similarity 99.1%; Pred. No. 6.1e-56;
Matches 108; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GDRRPLPVDRAGLKEKTLILLDYSTKNPVRTVNENFLSLQLDPSIIHDGWLDFLSSKRL 60
|||||
Db 42 gdrplpvdraaglkectlllldvstknprtvtvnenflslqldpsilhdgwlflsskrl 101
|||||
Qy 61 VTLARGLSPAFLRGGKRTDFLQFQNLNPNPAKSRGGPGPDYLLKNYEDD 109
|||||
Db 102 vtlarglspafirggkrtdfqlqnlrnpaksggpgpdyllknyede 150
|||||

RESULT 12
AAY97633
ID AAY97633 standard; Protein: 538 AA.
XX
AC AAY97633;
XX
DT 20-APR-2001 (first entry)
XX
DE Human heparanase, hnhp1 pn9 form, protein sequence.
XX
KW Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;
KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
KW gene therapy; human.
XX
OS Homo sapiens.
XX
PN WO200100643-A2.
XX
PD 04-JAN-2001.
XX
PF 19-JUN-2000; 2000WO-IL00358.
XX
PR 25-JUN-1999; 99US-0140801.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI Pecker I, Michal I, Itzhaki H;
XX
DR WPI; 2001-137930/14.
XX
DR N-PSDB; AAA91099.
XX
XX New polynucleotides and polypeptides that are distantly homologous to
PT heparanase, useful in wound healing, as well as in gene therapy
PT protocols for angiogenesis, restenosis, atherosclerosis, or
PT inflammation -
XX
PS Claim 10; Page 63; 67pp; English.
XX
CC This sequence represents a heparanase of the invention.
CC The heparanase DNA and protein sequences are useful in wound healing,
CC angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary
CC diseases, neurodegenerative diseases (such as Scrape, Alzheimer's
CC disease, and Creutzfeldt-Jakob disease) or viral infections. The
CC heparanase coding sequence is particularly useful in gene therapy.
XX
SQ Sequence 480 AA;

Query Match 91.1%; Score 566; DB 22; Length 480;
Best Local Similarity 99.1%; Pred. No. 6.1e-56;
Matches 108; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GDRRPLPVDRAGLKEKTLILLDYSTKNPVRTVNENFLSLQLDPSIIHDGWLDFLSSKRL 60
|||||
Db 42 gdrplpvdraaglkectlllldvstknprtvtvnenflslqldpsilhdgwlflsskrl 101
|||||
Qy 61 VTLARGLSPAFLRGGKRTDFLQFQNLNPNPAKSRGGPGPDYLLKNYEDD 109
|||||
Db 102 vtlarglspafirggkrtdfqlqnlrnpaksggpgpdyllknyede 150
|||||

RESULT 13
AAB08851
ID AAB08851 standard; Protein: 535 AA.
XX
AC AAB08851;
XX
DT 15-JAN-2001 (first entry)
XX
DE A murine heparanase polypeptide.
XX
KW Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
KW heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW wound healing; infection; burn; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease;
KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease.
XX
OS Mus sp.
XX
PN WO200052178-A1.
XX
PD 08-SEP-2000.
XX
PF 14-FEB-2000; 2000WO-US03542.
XX
PR 01-MAR-1999; 99US-0258892.
XX

FT
XX
PN WO200100643-A2.
XX
PD 04-JAN-2001.
XX
PF 19-JUN-2000; 2000WO-IL00358.
XX
PR 25-JUN-1999; 99US-0140801.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI Pecker I, Michal I, Itzhaki H;
XX
DR WPI; 2001-137930/14.
XX
DR N-PSDB; AAA91099.
XX
XX New polynucleotides and polypeptides that are distantly homologous to
PT heparanase, useful in wound healing, as well as in gene therapy
PT protocols for angiogenesis, restenosis, atherosclerosis, or
PT inflammation -
XX
PS Claim 10; Page 61-62; 67pp; English.
XX
CC This sequence represents a heparanase of the invention.
CC The heparanase DNA and protein sequences are useful in wound healing,
CC angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary
CC diseases, neurodegenerative diseases (such as Scrape, Alzheimer's
CC disease, and Creutzfeldt-Jakob disease) or viral infections. The
CC heparanase coding sequence is particularly useful in gene therapy.
XX
SQ Sequence 538 AA;

Query Match 91.1%; Score 565.5; DB 22; Length 538;
Best Local Similarity 95.0%; Pred. No. 8.1e-56;
Matches 113; Conservative 0; Mismatches 1; Indels 5; Gaps 2;

Qy 1 GDRRPLPVDRAGLKEKTLILLDYSTKNPVRTVNENFLSLQLDPSIIHDGWLDFLSSKRL 60
|||||
Db 42 gdrplpvdraaglkectlllldvstknprtvtvnenflslqldpsilhdgwlflsskrl 101
|||||
Qy 61 VTLARGLSPAFLRGGKRTDFLQFQNLNPNPAKSRGGPGPDYLLKNYEDD 119
|||||
Db 102 vtlarglspafirggkrtdfqlqnlrnpaksggpgpdyllknyed--ars---ldk 155
|||||

RESULT 13
AAB08851
ID AAB08851 standard; Protein: 535 AA.
XX
AC AAB08851;
XX
DT 15-JAN-2001 (first entry)
XX
DE A murine heparanase polypeptide.
XX
KW Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
KW heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW wound healing; infection; burn; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease;
KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease.
XX
OS Mus sp.
XX
PN WO200052178-A1.
XX
PD 08-SEP-2000.
XX
PF 14-FEB-2000; 2000WO-US03542.
XX
PR 01-MAR-1999; 99US-0258892.
XX

XX	Polynucleotides encoding mammalian endoglucuronidases, especially
PT	heparanases, useful to promote wound healing
XX	
PS	Claim 6; Page 69-73; 112pp; English.
XX	
CC	The invention relates to nucleic acid sequences that encode heparanase
CC	enzymes having endoglucuronidase activity. Recombinant heparanases are
CC	capable of removing the HS side chain from heparan sulfate proteoglycan
CC	(HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to
CC	inhibit heparanase, this is useful for treatment of a physiological or
CC	medical condition associated with elevated heparanase activity, such as
CC	metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,
CC	arteriosclerosis, atherosclerosis and inflammation. The human, murine and
CC	rat heparanases can be used to enhance wound healing, especially
CC	associated with tissue development and repair. The conditions mentioned
CC	above can be diagnosed using specific antibodies, and also using primers
CC	and probes specific for the heparanase polynucleotides. Other uses of the
CC	heparanases include sequencing sulfated molecules such as HSPG. The
CC	present sequence represents a human heparanase.
XX	
SQ	Sequence 543 AA;
XX	

```

Query Match      25.2%; Score 156.5; DB 20; Length 543;
Best Local Similarity 37.2%; Pred. No. 1.2e-09;
Matches 42; Conservative 20; Mismatches 44; Indels 7; Gaps 3;

QY      1  GDRRLPLVDRAAGLUKXTLLLDVSTNPKVTNNENFLSLQLDPSIIHD-GWLDFLSKRR 59
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      22  gp1spgpalrpa--qaqdvvdldfftqphlvpsflsvtldanlatdprfllllgspk 79

QY      60  LVTLARGLSPAFLLPFGCKRTDFLQFQMLRNPAPKSRGGGPPDYILKNTEDDIVR 112
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      80  lrtlarglspaylrfgtktdif-----dpkkestfeersywgsgvnaqdlck 128

```

Search completed: July 30, 2002, 08:14:02
Job time: 339 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:16:27 ; Search time 71.68 Seconds
(without alignments)
160.864 Million cell updates/sec

Title: US-09-836-461-2_COPY_42_161

Perfect score: 621

Sequence: 1 GDRRLPVDRAAGLKEKTLI.....YVLKNEYEDIVSDVALDKQ 120

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	566	91.1	480	JC7506	heparanase protein
2	85.5	13.8	1391	T20642	hypothetical prote
3	85.5	13.8	1397	F87998	protein F09C3.1 li
4	84.5	13.6	334	A86397	protein T7N9.6 [im
5	77.5	12.5	481	S56299	hypothetical prote
6	76	12.2	332	A81675	conserved hypothet
7	75	12.1	329	E71516	hypothetical prote
8	74	11.9	150	T34697	probable insertion
9	73.5	11.8	253	AF3625	call division inh
10	73.5	11.8	581	T12095	beta-fructofuranos
11	72	11.6	160	T37000	probable transpos
12	72	11.6	293	T09171	ribosomal protein
13	72	11.6	293	T09170	ribosomal protein
14	72	11.6	582	A84206	hypothetical prote
15	71.5	11.5	1283	S52500	SWH1 protein homol
16	70.5	11.4	188	AC3613	probable s-adenosy
17	70	11.3	1234	S52099	phospholipase C be
18	70	11.3	1234	T38994	phospholipase C-be
19	70	11.3	1983	T00385	KIAA0624 protein -
20	69	11.1	319	AC3523	aliphatic sulfonat
21	68.5	11.0	378	A81275	hypothetical prote
22	68.5	11.0	613	JC6033	mosquitocidal prot
23	68.5	11.0	694	F90609	exodeoxyribonuclea
24	68	11.0	900	C64232	alanine--trna liga
25	68	11.0	1084	B64088	hemoglobin-binding
26	67.5	10.9	165	T36981	probable insertion
27	67.5	10.9	492	S71146	H-transposon AT
28	67	10.8	599	S18735	centromere protein
29	67	10.8	742	S58691	kinesin-related pr

30	66	10.6	289	2	A12165	bicarbonate transp
31	66	10.6	367	2	JC2560	flagellin fliC-2 -
32	65.5	10.5	332	2	C95150	SAP domain protein
33	65.5	10.5	433	2	C82839	thymidine phosphor
34	65	10.5	327	2	T39981	rev interacting pr
35	65	10.5	652	2	D96808	protein F28K19.6 [
36	65	10.5	830	2	F83659	stage II sporulati
37	65	10.5	1299	2	AH2090	two-component hybr
38	65	10.5	2288	2	T30568	acetyl-CoA carboxy
39	64.5	10.4	251	1	XMEBFF	flagellar basal bo
40	64.5	10.4	251	2	AC0640	probable flagellar
41	64.5	10.4	296	2	A64035	hypothetical prote
42	64	10.3	274	2	G96799	hypothetical prote
43	64	10.3	332	2	S77386	nitrate transport
44	64	10.3	345	2	S59236	SAE2 protein - yea
45	64	10.3	363	2	S44443	RAD23 protein homo

ALIGNMENTS

RESULT 1

JC7506

heparanase protein 2a - human

C:Species: Homo sapiens (man)

C:Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 01-Dec-2000

C:Accession: JC7506

R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hirccock, M.;

Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000

A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase famli

A:Reference number: JC7506

A:Accession: JC7506

A:Molecule type: mRNA

A:Residues: 1-480 <MCK>

A:Cross-references: GB:AF282885

C:Comment: This protein, a intracellular membrane-bound enzyme, has biological and th

therapies.

C:Genetics:

A:Gene: hpa2a

A:Map position: 10q23-10q24

C:Keywords: heparin binding; membrane bound

Query Match 91.1%; Score 566; DB 2; Length 480;

Best Local Similarity 99.1%; Pred. No. 5.8e-53;

Matches 108; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GDRRLPVDRAAGLKEKTLIILLDVSTKNPVRTVNENFLSLQLDPSIIHDGWLDFLSSKRL 60

|||||

Db 42 GDRRLPVDRAAGLKEKTLIILLDVSTKNPVRTVNENFLSLQLDPSIIHDGWLDFLSSKRL 101

|||||

Oy 61 VTLARGLSPAFIRFGGKRTDFLQFONLRNPAKSRGGPGDYLYKNVEDD 109

|||||

Db 102 VTLARGLSPAFIRFGGKRTDFLQFONLRNPAKSRGGPGDYLYKNYEDE 150

|||||

RESULT 2

T20642

hypothetical protein F09C3.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T20642

R:Lennard, N.

submitted to the EMBL Data Library, March 1997

A:Reference number: Z19304

A:Accession: T20642

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1391 <N1L>

A:Cross-references: EMBL:Z92781; PIDN:CA807179.2; GSPDB:GN00019; CESP:F09C3.1

A:Experimental source: clone F09C3

C:Genetics:

A:Gene: CESP:F09C3.1

Db 334 KTVIPAKVFGFSIRTVPMDSSEKTSVLQKHCDAKFKSLNSPNKCRTELHIDGAYWSD 393
Qy 53 ----DFLSSKRLVTLARGSLPAFLRFGGKRTDFLQFQNLNPA-----KSRGGPG 98
Db 394 PFNAQFTAAKATKLVGVDPDFTRFGSGSIPITLTFQDALNTSVLLLPWGRGDDG 448
RESULT 6
A81675
conserved hypothetical protein TC0713 [imported] - Chlamydia muridarum (strain Nigg)
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: A81675
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: A81675
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-332 <TET>
A:Cross-references: GB:AE002339; GB:AE002160; NID:g7190728; PIDN:AAF39526.1; PID:g719074
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0713

Query Match 12.2%; Score 76; DB 2; Length 332;
Best Local Similarity 28.8%; Pred. No. 2.3;
Matches 34; Conservative 19; Mismatches 45; Indels 20; Gaps 6;
Qy 3 RRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHDGWLDFLSKRLVT 62
Db 202 KGRPPKQSAKVEAETVSSDIYTKVP--SAARRELF--PEITSASLTF--SEKFT 254
Qy 63 LARGSLPAFLRFGGKRTDFL-----QFQNLNPAKSRG-----GPGDPYILKNYEDD 109
Db 255 EEEFL--AHLRGGGRLEQNLAKFSERFDSLRELSAKILGYDGDGASDFGEEYDD 310

RESULT 7
E71516
hypothetical protein CT429 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession: E71516
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tra
A:Reference number: A71570; MUID:99000809
A:Accession: E71516
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-329 <ARN>
A:Cross-references: GB:AE001316; GB:AE001273; NID:g3328850; PIDN:AAC68026.1; PID:g332885
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: CT429

Query Match 12.1%; Score 75; DB 2; Length 329;
Best Local Similarity 29.7%; Pred. No. 2.9;
Matches 35; Conservative 19; Mismatches 44; Indels 20; Gaps 6;
Qy 3 RRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHDGWLDFLSKRLVT 62
Db 202 KGRPPKQSAKVEAETVSSDIYTKVP--SVARRELF--PEITSASLTF--SEKFT 254

Qy 63 LARGSLPAFLRFGGKRTDFL-----QFQNLNPAKSRG-----GPGDPYILKNYEDD 109
Db 255 EEEFL--AHLRGGGRLEQNLAKFSERFDSLRELSAKILGYDSDGETGDFDEEYDD 310

RESULT 8
T34697
probable insertion element ISL647 transposase, truncated - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T34697
R:Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, May 1998
A:Reference number: T34697
A:Accession: T34697
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-150 <OLI>
A:Cross-references: EMBL:AL023702; PIDN:CAAL9229.1; GSPDB:GN00070; SCOEEDB:SC1C3.05
A:Experimental source: strain A3(2)
C:Comment: The complete transposase is probably translated by ribosomal frameshifting
C:Genetics:
A:Gene: SCOEEDB:SC1C3.05

Query Match 11.9%; Score 74; DB 2; Length 150;
Best Local Similarity 29.9%; Pred. No. 1.4;
Matches 29; Conservative 9; Mismatches 33; Indels 26; Gaps 4;
Qy 5 PLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHDGWLDFLSKRLVTLA 64
Db 4 PNPTDR--GKKGSKIHLVDRRGLP-----LSIGISAANLHD-----SQALIPLV 46

Qy 65 RGLSPAFLRFGGKRTDFLQFQNLNPAKSRGGPGDPY 101
Db 47 RGIPPTIRSRGPR-----RRPGKLHGDKGYDY 74

RESULT 9
AF3625
cell division inhibitor minC [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C:Accession: AF3625
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Muler, C.; Los, T.; Ivanov
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Let
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit
A:Reference number: AB3252; PMID:11756688
A:Accession: AF3625
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-253 <KUR>
A:Cross-references: GB:AE008918; PIDN:AAL54169.1; PID:g17985135; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME110927
A:Map position: II
C:Superfamily: cell division inhibitor minC

Query Match 11.8%; Score 73.5; DB 2; Length 253;
Best Local Similarity 37.3%; Pred. No. 3;
Matches 22; Conservative 7; Mismatches 21; Indels 9; Gaps 2;
Qy 29 PVRTVNENFLSLQLDPSIIHDGWLDFLSKRLVTLARGSLPAFLRFGGKRTDFLQFQNL 87
Db 15 PIRLGRSFLAWVLSPELPGWLE-----RLDDLARRSSGFPL----GRPVLDMLNL 64

RESULT 10
T12095
beta-fructofuranosidase (EC 3.2.1.26), cell wall - fava bean
N:Alternate names: cell wall invertase II
C:Species: Vicia faba (fava bean)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000

RESULT 12

T09171
ribosomal protein L4 - spinach
C:Species: Spinacia oleracea (spinach)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 20-Jun-2000
C:Accession: T09171
R:Trifa, Y.; Privat, I.; Cagnon, J.; Baeza, L.; Lerbs-Mache, S.
J. Biol. Chem. 273, 3980-1985, 1998
A:Title: The nuclear RPL4 gene encodes a chloroplast protein that co-purifies with the

```

Query Match      11.6%; Score 72; DB 2; Length 293;
Best Local Similarity 29.6%; Pred. No. 5.3;
Matches 29; Conservative 14; Mismatches 31; Indels 24; Gaps 5
QY      5 PLPVDRAAGLKEKTLILLDVTSTKNPVRTVNEFLSLQLDP-----SIITHGMLDPLFSKR 5
      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      54 PLPLNFSSEK-----VCETFLNLKTAPEKARAVVHRLIHLTHLONKR 9
QY      60 IYTLARGLSPAFRLPGGKRTDLQFONLRNPAKSRGGP 97
      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 14
AB4206
hypothetical protein porA [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: AB4206
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla
; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddo

Db 286 SSK 288

us-09-836-461-2_copy_42_161.rpr

Tue Jul 30 10:12:03 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:31:30 ; Search time 36.14 Seconds
(without alignments)
128.565 Million cell updates/sec

Title: US-09-836-461-2_COPY_42_161
Perfect score: 621
Sequence: 1 GORRPLVDRAGLKEKTLI.....YLLKNYEDDIVRSDVALDKQ 120

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	77.5	12.5	481	1	YFL4_YEAST
2	76	12.2	332	1	W13_CHLMU
3	75	12.1	329	1	Y429_CHLTR
4	72	11.6	293	1	RK4_SPIOL
5	71	11.4	897	1	SVI_FERPE
6	70.5	11.4	992	1	DLP4_RAT
7	70	11.3	1234	1	PIP3_HUMAN
8	68.5	11.0	613	1	CGAA_CLOB1
9	68	11.0	900	1	SV4_MYCGE
10	68	11.0	1067	1	HGBB_HAEIN
11	68	11.0	1084	1	HGP3_HAEIN
12	67	10.8	599	1	CENB_HUMAN
13	67	10.8	742	1	K121_STRPU
14	66	10.6	367	1	FLC2_PROMI
15	66	10.6	599	1	CENB_MOUSE
16	64.5	10.4	251	1	FLGF_SALTY
17	64.5	10.4	296	1	YF23_HAEIN
18	64.5	10.4	989	1	DLP4_HUMAN
19	64	10.3	274	1	TL30_ARATH
20	64	10.3	345	1	SAE2_YEAST
21	64	10.3	363	1	R23A_HUMAN
22	64	10.3	598	1	HEXA_ALTSO
23	64	10.3	750	1	PSAA_MESVI
24	63.5	10.2	491	1	CAT2_MAIZE
25	63.5	10.2	492	1	ATPB_ANTFO
26	63	10.1	219	1	C270_MOUSE
27	63	10.1	372	1	3BH3_MOUSE
28	63	10.1	671	1	VINE_HUMAN
29	63	10.1	987	1	SOXA_RHIME
30	62.5	10.1	476	1	ATPB_GALSU
31	62.5	10.1	498	1	ATPB_NICPL
32	62.5	10.1	498	1	ATPB_NICRP
33	62.5	10.1	498	1	ATPB_NICSP

34	62.5	10.1	498	1	ATPB_TORAC
35	62.5	10.1	575	1	UL87_EBV
36	62.5	10.1	732	1	ACET_HUMAN
37	62.5	10.1	1054	1	POL_SIVMK
38	62.5	10.1	1118	1	UBP8_HUMAN
39	62.5	10.1	1306	1	ACE_HUMAN
40	62	10.0	363	1	R23A_MOUSE
41	62	10.0	369	1	CD5S_MOUSE
42	62	10.0	382	1	TP6A_PYRHO
43	62	10.0	606	1	CENB_CRIGR
44	61.5	9.9	299	1	LAXC_LACLA
45	61.5	9.9	299	1	LAXP_LACLA

ALIGNMENTS

RESULT 1					
YFL4_YEAST					
ID	YFL4_YEAST	STANDARD;	PRT;	481 AA.	
AC	P43616;				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	01-OCT-1996 (Rel. 34, Last annotation update)				
DE	Hypothetical 52.9 kDa protein in SAPI55-YMR31 intergenic region.				
GN	YFR044C.				
OS	Saccharomyces cerevisiae (Baker's Yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.				
OX	NCBI_TaxID=4932;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=S288C / AB972;				
RX	MEDLINE=95400292; PubMed=7670463;				
RA	Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,				
RA	Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,				
RA	Yamazaki M., Tashiro H., Eki T.;				
RT	"Analysis of the nucleotide sequence of chromosome VI from				
RT	Saccharomyces cerevisiae."				
RL	Nat. Genet. 10:261-268(1995).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=S288C / AB972;				
RX	MEDLINE=96287652; PubMed=8686379;				
RA	Eki T., Naitou M., Hagiwara H., Ozawa M., Sasanuma S.-I.,				
RA	Sasanuma M., Tsuchiya Y., Shibata T., Hanaoka F., Murakami Y.;				
RT	"Analysis of a 36.2 kb DNA sequence including the right telomere of				
RT	chromosome VI from Saccharomyces cerevisiae."				
RL	Yeast 12:149-167(1996).				
CC	-!- SIMILARITY: STRONG, TO THE C-TERMINAL HALF OF YEAST YBR281C.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; D50617; BAA09283.1; -				
DR	EMBL; D44597; BAA08010.1; -				
DR	SGD; S0001940; YFR044C.				
DR	InterPro; IPR002933; Peptidase_M20.				
DR	Pfam; PF01546; Peptidase_M20; 1.				
KW	Hypothetical protein.				
SQ	SEQUENCE 481 AA; 52871 MW; 3E53773A945F5EBC CRC64;				

Query Match 12.5%; Score 77.5; DB 1; Length 481;
Best Local Similarity 27.8%; Pred No. 0.98;
Matches 32; Conservative 12; Mismatches 38; Indels 33; Gaps 5;
QY 17 KTLILLDVSTKNPVRTV-----NENFLSL-----QLDPSIIHDG--WL-- 52

DB 334 KTVIPAKVGFSGSIRTPDMDSEKLTSLVQKHDAKFKSLNPNKCRTELLHDGAYWSD 393
QY 53 ----DFLSKRLVTLARGLSAPFLRFGGKRTDFLQFQNLNPA-----KSRGGFG 98
DB 394 PNAOFTAAKATKLVYGVDPDFREGGSIPITLTFQDALNTSVLLLPMDGRGDDG 448
RESULT 2
Y713_CHLMU
ID Y713_CHLMU STANDARD; PRT; 332 AA.
AC Q9PW3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein TC0713.
GN TC0713
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / N199;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., Baas S.,
White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bowman C.,
Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salsberg S.L.,
Eisen J., Fraser C.M.;
RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39";
RT Nucleic Acids Res. 28:1397-1406(2000).
RL NCBI_TaxID=83560;
CC -!- SIMILARITY: BELONGS TO THE UPF0158 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AE002339; AAF39526.1; --
DR TIGR; TC0713;
KW Hypothetical protein; Complete proteome.
FT DOMAIN 301 316 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 308 316 POLY-ASP.
SQ SEQUENCE 332 AA; 39202 MW; 27527DC5E6ABDEA4 CRC64;
Query Match 12.2%; Score 76; DB 1; Length 332;
Best Local Similarity 28.8%; Pred. No. 0.89; Indels 20; Gaps 6;
Matches 34; Conservative 19; Mismatches 45
QY 3 RRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNEFLSLQDPSIHDGWLDFLSSKRLVT 62
DB 202 KRGPRPKQSAKVEATTSSDIYTKVP--SAARREFLF---PEITSASSLTF--SEKEDT 254
QY 63 LARGLSAPFLRFGGKRTDFL-----QFONLRNPAKSRG-----GPGPDYLNKYEDD 109
DB 255 EEEFL--AHLRGGGRLEQDLNLAKEFSERFDSRLSALGVDGDSGDFGDFEYDD 310
RESULT 3
Y429_CHLTR
ID Y429_CHLTR STANDARD; PRT; 329 AA.
AC O84436;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein CT429.
GN CT429.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
Mitchell W.P., Ollinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
Chlamydia trachomatis";
RL Science 282:754-759(1998).
CC -!- SIMILARITY: BELONGS TO THE UPF0158 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001316; AAC68026.1; --
DR Hypothetical protein; Complete proteome.
KW DOMAIN 301 313 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 301 313 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 329 AA; 39263 MW; B1706123BE7F7698 CRC64;
Query Match 12.1%; Score 75; DB 1; Length 329;
Best Local Similarity 29.7%; Pred. No. 1.1; Indels 20; Gaps 6;
Matches 35; Conservative 19; Mismatches 44
QY 3 RRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNEFLSLQDPSIHDGWLDFLSSKRLVT 62
DB 202 KRGPRPKQSAKVEATTSSDIYTKVP--SVARREFLF---PEITSPSSLTF--SEKEDT 254
QY 63 LARGLSAPFLRFGGKRTDFL-----QFONLRNPAKSRG-----GPGPDYLNKYEDD 109
DB 255 EEEFL--AHLRGGGRLEQDLNLAKEFSERFDSRLSALGVDGDSGDFGDFEYDD 310
RESULT 4
RK4_SPIOL
ID RK4_SPIOL STANDARD; PRT; 293 AA.
AC O49937; O49938;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L4, chloroplast precursor (R-protein L4).
GN RPL4.
OS Spinacia oleracea (Spinach).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllaceae; Chenopodiaceae; Spinacia.
OX NCBI_TaxID=3562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 51-63.
RC STRAIN=CV, MELODY; TISSUE=Leaf;
RX MEDLINE=98129805; PubMed=9461586;
RA Trifa V., Privat I., Gagnon J., Baeza L., Lerbs-Mache S.;
RT "The nuclear RPL4 gene encodes a chloroplast protein that co-purifies
with the 70S-like transcription complex as well as plastid
ribosomes";
RL J. Biol. Chem. 273:3980-3985(1998).
RN [2]
RP SEQUENCE OF 51-68; 82-86; 98-105; 131-137; 152-158; 211-224 AND
RP 228-235.
RC STRAIN=CV, ALVARO; TISSUE=Leaf;
RX MEDLINE=20435798; PubMed=10874046;
RA Yamaguchi K., Subramanian A.R.;
RT "The plastid ribosomal proteins. Identification of all the proteins in
the 50 S subunit of an organelle ribosome (chloroplast)";
RL J. Biol. Chem. 275:28466-28482(2000).

DB 334 KTVIPAKVGFSGSIRTPDMDSEKLTSLVQKHDAKFKSLNPNKCRTELLHDGAYWSD 393
QY 53 ----DFLSKRLVTLARGLSAPFLRFGGKRTDFLQFQNLNPA-----KSRGGFG 98
DB 394 PNAOFTAAKATKLVYGVDPDFREGGSIPITLTFQDALNTSVLLLPMDGRGDDG 448
RESULT 2
Y713_CHLMU
ID Y713_CHLMU STANDARD; PRT; 332 AA.
AC Q9PW3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein TC0713.
GN TC0713
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / N199;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., Baas S.,
White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bowman C.,
Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salsberg S.L.,
Eisen J., Fraser C.M.;
RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39";
RT Nucleic Acids Res. 28:1397-1406(2000).
RL NCBI_TaxID=83560;
CC -!- SIMILARITY: BELONGS TO THE UPF0158 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AE002339; AAF39526.1; --
DR TIGR; TC0713;
KW Hypothetical protein; Complete proteome.
FT DOMAIN 301 316 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 308 316 POLY-ASP.
SQ SEQUENCE 332 AA; 39202 MW; 27527DC5E6ABDEA4 CRC64;
Query Match 12.2%; Score 76; DB 1; Length 332;
Best Local Similarity 28.8%; Pred. No. 0.89; Indels 20; Gaps 6;
Matches 34; Conservative 19; Mismatches 45
QY 3 RRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNEFLSLQDPSIHDGWLDFLSSKRLVT 62
DB 202 KRGPRPKQSAKVEATTSSDIYTKVP--SAARREFLF---PEITSASSLTF--SEKEDT 254
QY 63 LARGLSAPFLRFGGKRTDFL-----QFONLRNPAKSRG-----GPGPDYLNKYEDD 109
DB 255 EEEFL--AHLRGGGRLEQDLNLAKEFSERFDSRLSALGVDGDSGDFGDFEYDD 310
RESULT 3
Y429_CHLTR
ID Y429_CHLTR STANDARD; PRT; 329 AA.
AC O84436;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein CT429.
GN CT429.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
Mitchell W.P., Ollinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
Chlamydia trachomatis";
RL Science 282:754-759(1998).
CC -!- SIMILARITY: BELONGS TO THE UPF0158 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001316; AAC68026.1; --
DR Hypothetical protein; Complete proteome.
KW DOMAIN 301 313 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 301 313 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 329 AA; 39263 MW; B1706123BE7F7698 CRC64;
Query Match 12.1%; Score 75; DB 1; Length 329;
Best Local Similarity 29.7%; Pred. No. 1.1; Indels 20; Gaps 6;
Matches 35; Conservative 19; Mismatches 44
QY 3 RRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNEFLSLQDPSIHDGWLDFLSSKRLVT 62
DB 202 KRGPRPKQSAKVEATTSSDIYTKVP--SVARREFLF---PEITSPSSLTF--SEKEDT 254
QY 63 LARGLSAPFLRFGGKRTDFL-----QFONLRNPAKSRG-----GPGPDYLNKYEDD 109
DB 255 EEEFL--AHLRGGGRLEQDLNLAKEFSERFDSRLSALGVDGDSGDFGDFEYDD 310
RESULT 4
RK4_SPIOL
ID RK4_SPIOL STANDARD; PRT; 293 AA.
AC O49937; O49938;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L4, chloroplast precursor (R-protein L4).
GN RPL4.
OS Spinacia oleracea (Spinach).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllaceae; Chenopodiaceae; Spinacia.
OX NCBI_TaxID=3562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 51-63.
RC STRAIN=CV, MELODY; TISSUE=Leaf;
RX MEDLINE=98129805; PubMed=9461586;
RA Trifa V., Privat I., Gagnon J., Baeza L., Lerbs-Mache S.;
RT "The nuclear RPL4 gene encodes a chloroplast protein that co-purifies
with the 70S-like transcription complex as well as plastid
ribosomes";
RL J. Biol. Chem. 273:3980-3985(1998).
RN [2]
RP SEQUENCE OF 51-68; 82-86; 98-105; 131-137; 152-158; 211-224 AND
RP 228-235.
RC STRAIN=CV, ALVARO; TISSUE=Leaf;
RX MEDLINE=20435798; PubMed=10874046;
RA Yamaguchi K., Subramanian A.R.;
RT "The plastid ribosomal proteins. Identification of all the proteins in
the 50 S subunit of an organelle ribosome (chloroplast)";
RL J. Biol. Chem. 275:28466-28482(2000).


```

Db    711 YKQNEFDNL 719      ! :!::!!:

RESULT 11
HGP3_HAEIN STANDARD; PRT; 1084 AA.
AC P44836;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-WAR-2002 (Rel. 41, Last annotation update)
DE Probable hemoglobin and hemoglobin-haptoglobin binding protein 3
DE precursor.
DE HI0712.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus
OX NCBI_Taxid=727;
[1]
SEQUENCE FROM N.A.
STRAIN-RD / KW20 / ATCC 51907;
MEDLINE=95350630; PubMed=7542800;
Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fritchman J.L., Glodek A., Kelley J.M.,
RA Scott J.D., Shirley R., Liu L.-I., Fields C.A. Cocayne J.D.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
[2]
IDENTIFICATION BY MASS SPECTROMETRY.
RP MEDLINE=20137488; PubMed=10675023;
RX Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
RL Electrophoresis 21:414-429(2000).
CC -1- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE
CC HEMOGLOBIN/HAPTOGLOBIN COMPLEX OF THE HUMAN HOST AND IS REQUIRED
CC FOR HEME UPTAKE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCAA
CC REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAIRING.
CC ADDITION OR LOSS OF CCAA REPEAT UNITS WOULD CHANGE THE READING
CC FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
CC REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A WAY TO
CC AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR FAMILY;
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
-----
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U32754; AAC22369.1; -
DR TIGR; HI0712; -
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
DR Outer membrane; Transport; TonB box; Multigene family; Signal;
KW Receptor; Repeat; Complete proteome.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 1084 PROBABLE HEMOGLOBIN AND HEMOGLOBIN-
```

FT DOMAIN 26 73 HAPTOGLOBIN BINDING PROTEIN 3.
 FT REPEAT 26 29 12 X 4 AA TANDEM REPEATS OF Q-P-T-N.
 FT REPEAT 30 33 1.
 FT REPEAT 34 37 2.
 FT REPEAT 38 41 3.
 FT REPEAT 42 45 4.
 FT REPEAT 46 49 5.
 FT REPEAT 50 53 6.
 FT REPEAT 54 57 7.
 FT REPEAT 58 61 8.
 FT REPEAT 62 65 9.
 FT REPEAT 66 69 10.
 FT REPEAT 70 73 11.
 FT SITE 83 90 TONB BOX.
 FT SITE 1067 1084 TONB C-TERMINAL BOX.
 FT SITE 1067 1084 TONB C-TERMINAL BOX.
 FT SEQUENCE 1084 AA; 123955 MW; 794DF91E0F53CFD9 CRC64;

Query Match 11.0%; Score 68; DB 1; Length 1084;
 Best Local Similarity 26.9%; Pred. No. 28;
 Matches 35; Conservative 16; Mismatches 47; Indels 32; Gaps 8;

QY 1 GDRPLP-VDRAAGL-----REKTLILLDYSTKNPVRTVNN-----FLSLQIDPSIIHDG 50
 Db 619 GKPTAPDPYSHRCSLMNTDQKETYLIPTVTKNNVLYFGDNVQLTSWLGDLNRYDHVK 678
 QY 51 WLDFLSKRLVTLARGL-SPAFLRFGKGTDFLQFNLRNPAKSRGGPGDY----- 101
 Db 679 YLPSYDEK--IPVNGLLTGLFKKFGPK--DIVYGSYSKPKA-----DYTCTYNSD 726
 QY 102 -YLNKYEDDI 110
 Db 727 CYKKNFKDNL 736

RESULT 12
 CENB_HUMAN STANDARD; PRT; 599 AA.
 AC P07199;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Major centromere autoantigen B (Centromere protein B) (CENP-B).
 GN CENPB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91372020; PubMed=1893793;
 RA Sullivan K.F., Glass C.A.;
 RT "CENP-B is a highly conserved mammalian centromere protein with
 RT homology to the helix-loop-helix family of proteins.";
 RL Chromosoma 100:360-370(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggeley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhami P., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Leinvaslao M.H., Leversha M., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,

RA Milne S., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RA "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 RN [3]
 RP SEQUENCE OF 6-599 FROM N.A.
 RX MEDLINE=87166180; PubMed=2435739;
 RA Earnshaw W.C., Sullivan K.F., Machlin P.S., Cooke C.A.,
 RA Kaiser D.A., Pollard T.D., Rothfield N.F., Cleveland D.W.;
 RT "Molecular cloning of cDNA for CENP-B, the major human centromere
 RT autoantigen.";
 RL J. Cell Biol. 104:817-829(1987).
 RN [4]
 RP SUBUNITS, AND DOMAINS.
 RX MEDLINE=93107144; PubMed=1469042;
 RA Yoda K., Kitagawa K., Masumoto H., Muro Y., Okazaki T.;
 RT "A human centromere protein, CENP-B, has a DNA binding domain
 RT containing four potential alpha helices at the NH2 terminus, which is
 RT separable from dimerizing activity.";
 RL J. Cell Biol. 119:1413-1427(1992).
 RN [5]
 RP STRUCTURE BY NMR OF 1-56.
 RX MEDLINE=98119825; PubMed=9451007;
 RA Iwahara J., Kigawa T., Kitagawa K., Masumoto H., Okazaki T.,
 RA Yokoyama S.;
 RT "A helix-turn-helix structure unit in human centromere protein B
 RT (CENP-B).";
 RL EMBO J. 17:827-837(1998).
 CC -!- FUNCTION: INTERACTS WITH CENTROMERIC HETEROCHROMATIN IN
 CC CHROMOSOMES AND BINDS TO A SPECIFIC SUBSET OF ALPHOID SATELLITE
 CC DNA, CALLED THE CENP-B BOX. MAY ORGANIZE ARRAYS OF CENTROMERE
 CC SATELLITE DNA INTO A HIGHER ORDER STRUCTURE WHICH THEN DIRECTS
 CC CENTROMERE FORMATION AND KINETOCORE ASSEMBLY IN MAMMALIAN
 CC CHROMOSOMES.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: STRONG, WITH OTHER MAMMALIAN CENP-B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X55039; CAA38879.1; -;
 DR EMBL; AL109804; CAC17547.1; -;
 DR EMBL; X05299; CAA28918.1; -;
 DR PIR; A27272; A27272.
 DR PIR; S18735; S18735.
 DR PDB; 1BW6; 07-OCT-98.
 DR MIM; 117140; -;
 KW Chromosomal protein; Nuclear protein; DNA-binding; Centromere;
 KW 3D-structure.
 FT DNA_BIND 1 125
 FT DOMAIN 404 465 GLU-RICH (ACIDIC)
 FT DOMAIN 508 538 ASP/GLU-RICH (ACIDIC).
 FT CONFLICT 583 583 R -> M (IN REF. 3).
 FT CONFLICT 592 593 VR -> LL (IN REF. 3).
 SQ SEQUENCE 599 AA; 65171 MW; 9B4B7DB957A914AA CRC64;

Query Match 10.8%; Score 67; DB 1; Length 599;
 Best Local Similarity 28.7%; Pred. No. 17;

Db	301	QDSLGGNAKTVVMWNGPASYNEDEITITLRYANRAKNIKPKINEDP-KDALLREFQE	359
Qy	109	DIVRSVDALDKQ	120
Db	360	EISRLQALDKK	371

```
RESULT 14
FLC2_PROMI
ID FLC2
```

```

Query Match      10.6%; Score 66; DB 1; Length 367;
Best Local Similarity 26.9%; Pred. No. 12;
Matches 21; Conservative 13; Mismatches 32; Indels 12; Gaps 2;

43 DPSIIHGDWLFSLSKRLVTLARGLSPAFLRFGGKRTDFLOFQNLNRNPAKSGGPGDYV 102
204 NQVHNNQNNVAVKSGNFTVSGMTTCGEVETGSGKT-----TKFTADAGKDVK 272

```

RESULT	15
CENB_MOUSE	
ID	CENB_N
AC	P27790

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:30:45 ; Search time 123.94 Seconds
(without alignments)
167.495 Million cell updates/sec

Title: US-09-836-461-2_COPY_42_161
Perfect score: 621
Sequence: 1 GDRRLPVPDRAAGLKEKTLI.....YLLKNVEDDIVRSDVALDKQ 120

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	621	100.0	534	4 Q9HB38	Q9hb38 homo sapien
2	621	100.0	592	4 Q9HB37	Q9hb37 homo sapien
3	566	91.1	480	4 Q9HB39	Q9hb39 homo sapien
4	163	26.2	536	11 Q9QZF8	Q9qzf8 rattus norv
5	158.5	25.5	545	6 Q9MYX0	Q9myx0 bos taurus
6	157.5	25.4	545	4 Q9UL39	Q9ul39 homo sapien
7	156.5	25.2	543	4 Q9Y251	Q9y251 homo sapien
8	136.5	22.0	523	13 Q9QYK5	Q9qyk5 gallus gall
9	85.5	13.8	1391	5 O17772	O17772 caenorhabdi
10	84.5	13.6	304	10 Q9C5M5	Q9c5m5 arabidopsis
11	84.5	13.6	334	10 Q9LFY5	Q9lfy5 arabidopsis
12	82.5	13.3	756	16 Q988J0	Q988j0 rhizobium l
13	77	12.4	385	10 Q93X83	Q93x83 vitis labru
14	75	12.1	241	5 Q9W3U3	Q9w3u3 drosophila
15	74	11.9	150	2 Q98847	Q98847 streptomyce
16	73.5	11.8	581	10 Q43856	Q43856 vicia faba

17	73.5	11.8	705	4 Q9H9S9	Q9h9s9 homo sapien
18	72	11.6	160	2 Q9RI77	Q9ri77 streptomyce
19	72	11.6	582	17 Q9HR25	Q9hr25 halobacteri
20	71.5	11.5	527	10 Q9LRC8	Q9lrc8 scutellaria
21	71.5	11.5	1283	3 Q12451	Q12451 saccharomyc
22	71	11.4	881	5 Q9GR11	Q9gr11 paramecium
23	70.5	11.4	433	2 Q937S1	Q937s1 bruceella me
24	70	11.3	1983	4 Q9Y4D6	Q9y4d6 homo sapien
25	69.5	11.2	378	2 Q939J7	Q939j7 campylobact
26	69.5	11.2	1167	16 Q9A205	Q9a205 streptococc
27	69	11.1	116	4 Q9UH21	Q9uh21 homo sapien
28	69	11.1	116	4 Q9Y319	Q9y319 homo sapien
29	69	11.1	203	4 Q9NQF3	Q9nqf3 homo sapien
30	69	11.1	314	4 Q9H4I8	Q9h4i8 homo sapien
31	69	11.1	329	10 Q9FGI3	Q9fgi3 arabidopsis
32	69	11.1	363	12 Q93494	Q93494 okra yellow
33	69	11.1	409	15 Q90N44	Q90n44 human immun
34	69	11.1	624	10 Q9FGH9	Q9fgh9 arabidopsis
35	69	11.1	838	10 Q9FG37	Q9fg37 arabidopsis
36	69	11.1	993	2 Q48915	Q48915 microcystis
37	68.5	11.0	378	16 Q9PMY3	Q9pm3 campylobact
38	68.5	11.0	463	8 Q9MRQ8	Q9mrq8 ceratophyll
39	68.5	11.0	694	16 Q98PE1	Q98pel mycoplasma
40	68	11.0	321	15 Q9OKT0	Q9okt0 human immun
41	68	11.0	331	5 P91889	P91889 sepia offic
42	68	11.0	622	11 Q923C5	Q923c5 mus musculu
43	68	11.0	2126	2 Q9S1A8	Q9s1a8 microcystis
44	68	11.0	2126	2 Q93LR2	Q93lr2 microcystis
45	68	11.0	2126	2 Q93LRL	Q93lrl microcystis

ALIGNMENTS

RESULT 1

Q9HB38 ID Q9HB38 PRELIMINARY: PRT; 534 AA.

AC Q9HB38;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HEPARANASE-LIKE PROTEIN HPA2B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20483645; PubMed=11027606;
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,
RA Hircok M., Patel S., Barry E., Stuberfield C., Terrett J., Page M.;
RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian
RT Heparanase Family Member.";
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).
DR EMBL; AF282886; AAG23422.1; -;
SQ SEQUENCE 534 AA; 60063 MW; C3DE5E900CB338C4 CRC64;

Query Match 100.0%; Score 621; DB 4; Length 534;
Best Local Similarity 100.0%; Pred. No. 1.4e-59;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDRRLPVPDRAAGLKEKTLIILLDVSTKNPVRTVNEFNFLSLQLDPSIHDGWLDFLSKRL 60

DB 42 GDRRLPVPDRAAGLKEKTLIILLDVSTKNPVRTVNEFNFLSLQLDPSIHDGWLDFLSKRL 101

QY 61 VTLARGLSPAFLRFQKGRDTDFLQFQNLNRPAPKSRGGPGPDYLLKNYEDDIVRSDVALDKQ 120

DB 102 VTLARGLSPAFLRFQKGRDTDFLQFQNLNRPAPKSRGGPGPDYLLKNYEDDIVRSDVALDKQ 161

RESULT 2

Q9HB37

Q9QZF8		PRELIMINARY;	PRT;	536 AA.
AC	Q9QZF8;			
DT	01-MAY-2000 (TREMBlrel. 13, Created)			
DT	01-MAY-2000 (TREMBlrel. 13, Last sequence update)			
DT	01-DEC-2001 (TREMBlrel. 19, Last annotation update)			
DE	HEPARANASE.			
GN	HEP.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	Podyma K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.,			
RA	"Heparanase from parathyroid cell line."			
RL	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF184967; AAF04563.1; -. 6208B1FD9EE28421 CRC64;			
SQ	SEQUENCE 536 AA; 60569 MW; 6208B1FD9EE28421 CRC64;			
Query Match 26.2%; Score 163; DB 11; Length 536;				
Best Local Similarity 39.7%; Pred. No. 2.2e-09;				
Matches 46; Conservative 17; Mismatches 47; Indels 6; Gaps 3;				
QY	1 GDRRLPVDRAAG-LKEKTLILLDVSTKNPVRTVNVNENFLSLQLDPSIIHD-GWLDLSSK 58			
Db				
QY	12 GRLRALTGCTPGAPTAKDVVDLEFYTKRLQFSVSPSLTITIDASLATDPRFTFLGSP 71			
Db				
QY	59 RLVTARGLSPAFLRFGGKRTDFLOFQNLNPAKSRGGPGDPYLLKNYEDDIIVRSD 114			
Db				
QY	72 RLRLARGLSPAFLRFGGKRTDFLIF---DPNKEPTSEERSYQSDNNDCGSE 123			
Db				
RESULT 5				
Q9MY00	PRELIMINARY;	PRT;	545 AA.	
AC	Q9MY00;			
DT	01-OCT-2000 (TREMBlrel. 15, Created)			
DT	01-JUN-2001 (TREMBlrel. 17, Last sequence update)			
DT	01-JUN-2001 (TREMBlrel. 17, Last annotation update)			
DE	HEPARANASE.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=PLACENTA;			
RA	Kizaki K., Nakano H., Takahashi T., Imal K., Hashizume K.;			
RT	"Expression of Heparanase mRNA in Bovine Placenta During Gestation."			
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF281160; AAF87301.2; -. 61076 MW; FAC4BDFD855B933 CRC64;			
SQ	SEQUENCE 545 AA; 61076 MW; FAC4BDFD855B933 CRC64;			
Query Match 25.5%; Score 158.5; DB 6; Length 545;				
Best Local Similarity 38.0%; Pred. No. 7e-09;				
Matches 41; Conservative 15; Mismatches 39; Indels 13; Gaps 3;				
QY	7 PVDRAAGLKEKTLILLDVSTKNPVRTVNVNENFLSLQLDPSIIHD-GWLDLSSKRLVTLAR 65			
Db				
QY	36 PADDAE-----LEFFTERPLHLVSPAFLSFTIDANLATDPRFTFLGSSKRLTLAR 87			
Db				
QY	66 GLSPAFLRFGGKRTDFLOFQNLNPAKSRGGPGDPYLLKNYEDDIIVRSD 113			
Db				
QY	88 GLAPALRFGGKRTDFLOFQNLNPAKSRGGPGDPYLLKNYEDDIIVRSD 131			
Db				
RESULT 6				
Q9UL39	PRELIMINARY;	PRT;	545 AA.	
ID	Q9UL39			

RC	TISSUE=PLACENTA;
RX	MEDLINE=99335379; PubMed=10405343;
RA	Kussle P.H., Hulmes J.D., Ludwig D., Patel S., Navarro E.C.,
RA	Seddon A.P., Giorgio N.A., Bohlen P.;
RT	"Cloning and Functional Expression of a Human Heparanase Gene.";
RL	Biochem. Biophys. Res. Commun. 261:183-187(1999).
DR	EMBL; AF165154; AAD45379.1; -
DR	EMBL; AF144325; AAD41342.1; -
DR	EMBL; AF155510; AAD54941.1; -
DR	EMBL; AF152376; AAD45669.1; -
QY	SEQUENCE 543 AA; 61176 MW; AD262EC267334AB2 CRC64;
QY	Query Match 25.2%; Score 156.5; DB 4; Length 543;
QY	Best Local Similarity 37.2%; Pred. No. 1.2e-08;
QY	Matches 42; Conservative 20; Mismatches 44; Indels 7; Gaps 3;
Dy	22 GPLSLGALPRPA--QAQDVVDLDFFTQEPLHLVSPSFLSTIDANLATDPRFLLIGSPK 79 : : : : :
QY	60 LVTIARGLSPAFLRFGCKRTDFLOFQNLRNPAKSRGGPGDYILKNVEDDIVR 112 : : : : :
Dy	80 LRTLARGLSPAYLRFGTKDTDLIF----DPKKESTFEERSYMQSQVNODICK 128 : : : : :
RESULT 8	
Q90YK5	PRELIMINARY; PRT; 523 AA.
ID Q90YK5	PRELIMINARY; PRT; 523 AA.
AC Q90YK5	
DT 01-DEC-2001	(TrEMBLrel. 19, Created)
DT 01-DEC-2001	(TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001	(TrEMBLrel. 19, Last annotation update)
DE	HEPARANASE.
OS	Gallus gallus (Chicken).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OX	Gallus.
NCBI_TaxID=9031;	
RN [1]	
RP	SEQUENCE FROM N.A.
RX	PubMed=11387326;
RA	Goldschmidt O., Zcharia E., Alingorn H., Guatta-Rangini Z., Atzmon R.,
RA	Michal I., Pecker I., Mitrani E., Vlodaysky I.;
RT	"Expression Pattern and Secretion of Human and Chicken Heparanase Are
RL	Determined by their Signal peptide Sequence.";
J. Biol. Chem. 276:29178-29187(2001).	
DR	EMBL; AY037007; AAK82648.1; -
QY	SEQUENCE 523 AA; 58386 MW; 8EB0B7B18C9BF881 CRC64;
QY	Query Match 22.0%; Score 136.5; DB 13; Length 523;
QY	Best Local Similarity 47.2%; Pred. No. 1.7e-06;
QY	Matches 34; Conservative 8; Mismatches 25; Indels 5; Gaps 2;
Dy	22 LDVSTKPNVRTNVNFSLQLDPSIIHD-GWLDPLSKRLVTLARGLSPAFLRFGGKRTD 80 : : : : : : : : :
Dy	21 LOLGLREPIGAVSPAFLSLTLDASLARDPREFVALLRKPKLHTLASGLSPGFRLFEGGTSTD 80 : : : : :
QY	81 FLOFQNLRNPAK 92
Dy	81 FLIF---NPVK 88
RESULT 9	
O17772	PRELIMINARY; PRT; 1391 AA.
ID O17772	PRELIMINARY; PRT; 1391 AA.
AC O17772	
DT 01-JAN-1998	(TrEMBLrel. 05, Created)
DT 01-MAY-2000	(TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001	(TrEMBLrel. 19, Last annotation update)
DE	F09C3.1 PROTEIN.
N	F09C3.1

OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
RP SEQUENCE FROM N.A.
RA Lennard N.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP Ecker J.R.;
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL: Z92781; CAB07179.2; -.
DR InterPro: IPR001715; Calponin_hom.
DR InterPro: IPR003247; CH_type.
DR InterPro: IPR000048; IQ.
DR InterPro: IPR001936; RasGAP.
DR InterPro: IPR000593; RasGAP_C.
DR Pfam: PF00612; IQ; 2.
DR ProDom: PD001527; CH_type; 1.
DR ProDom: PD008735; RasGAP_C; 1.
DR SMART: SM00033; CH; 1.
DR SMART: SM00015; IQ; 1.
DR PROSITE: PS50021; CH; 1.
DR PROSITE: PS50018; RAS_GTPASE_ACTIV_2; 1.
SQ SEQUENCE 1391 AA; 159210 MW; 794DF9EBA2E65AC CRC64;

Query Match 13.8%; Score 85.5; DB 5; Length 1391;
Best Local Similarity 23.8%; Pred. No. 2.2;
Matches 35; Conservative 15; Mismatches 52; Indels 45; Gaps 4;

QY 8 VDAAGLKKEKTLILLDVSTKN--VRTVNFNLSLQD----- 43
Db 485 VEAFLAEQVVRINKAGSEPKVRELEKLSFKLDGYREELLHWYAKRISEELDASKH 544

QY 44 PSIIHGWLDF-----LSSKRLVTLAGLSPAFIRFGKKTDFLOQNLNPAK 92
Db 545 PESSQDFWLDHKFPLGTIIYVNTSTKSHSTSQEPLHPLRNWNPLODFVAYENVNF--- 601

QY 93 SRGGPGDYLYKNVEDDIVRSDVALDK 119
Db 602 -----DEIWKKEKEVVKQALRK 621

RESULT 10
ID Q9C5M5 PRELIMINARY; PRT; 304 AA.
AC Q9C5M5; 2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 19, Last annotation update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 33.0 KDA PROTEIN (AT1G27000/T7N9.6).
GN T7N9.6.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full length cDNA of gene T7N9.6 (GI:8778952).";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

[2]
RN SEQUENCE FROM N.A.
RP Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RP Yamada K., Banh J., Banno F., Dale J.M., Goldsmith A.D., Lee J.M.,
RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Yamamura Y., Yu G.,
RA Yu S., Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C.,
RA Koesema E., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M.,
RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
RA Southwick A., Tracy S.E., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full length cDNA of gene T7N9.6 (GI:8778952).";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF360143; AAK25853.1; -.
DR EMBL: AY057525; AAL09765.1; -.
DR EMBL: AY056367; AAL07253.1; -.
KW Hypothetical protein.
SQ SEQUENCE 304 AA; 33013 MW; 463DB4579AF52195 CRC64;

Query Match 13.6%; Score 84.5; DB 10; Length 304;
Best Local Similarity 27.6%; Pred. No. 0.43;
Matches 34; Conservative 20; Mismatches 44; Indels 25; Gaps 5;

QY 10 RAAGLKKEKTLILLDVSTK--NPVRTVNFNLSLQDPSIIH-----DGWLDFLSSKRLV 61
Db 157 RTQNLDDKVEKQIDLSKEINSQVARENSISLEMDLESNLHNLITGLDGKDLTLEYKQDV 216

QY 62 TLARGSPAFIRFGKKTDFLOQNLNPAKSRG-----GPGPDYLYKNVEDDIV 111
Db 217 TNVFMNL-LYNYFGKSTKLPMEQQLQPVNQARNLLADVETKG-----LKNLAELF 269

QY 112 RSD 114
Db 270 KSN 272

RESULT 11
ID Q9LFY5 PRELIMINARY; PRT; 334 AA.
AC Q9LFY5; 2001 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE T7N9.6.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RA Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,
RA Shinn P., Altafi H., Bei Q., Chin C., Chlou J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Hwang B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsy N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC T7N9 from chromosome
I.";
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.


```
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DE CG12542 PROTEIN.
GN CG12542.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S.E., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Foaier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.B., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattai B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003439; AAP46220.1; -.
DR FlyBase; FBgn0029921; CG12542.
SQ SEQUENCE 241 AA; 27764 MW; 157F97682E80CA80 CRC64;

Query Match 12.1%; Score 75; DB 5; Length 241;
Best Local Similarity 29.9%; Pred. No. 3.6;
Matches 26; Conservative 15; Mismatches 40; Indels 6; Gaps 3;

QY 22 LDVSTKNPVRTVNFNLSQLDPS-----IHDGWLDFLSSKRLVTLARGLSPAFLRFGGK 77
| : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 49 LECTAQEAQVNT-GEQFTTYGDRSNTDLLVHGFDPNNTKDYVNIIRVGLSLTD-ALAAK 106
| : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |

QY 78 RTDFLQFONLRNPAKSRGGPGDYLYK 104
| : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 107 RASILDKLNIRHTRALRVLPAPDFISK 133
| : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |

RESULT 15
O69847 PRELIMINARY; PRT; 150 AA.
ID O69847
AC O69847;
```

```
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE PUTATIVE INSERTION ELEMENT ISI1647 TRANSPOSASE (PROBABLE INSERTION
DE ELEMENT ISI1647 TRANSPOSASE) (FRAGMENT).
GN SCIC3.05 OR SCIB2.18.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RC SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RC SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmid and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RT Mol. Microbiol. 21:77-96(1996).
RN [4]
RC SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RC SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL023702; CAAL9229.1; -.
DR EMBL; AL356812; CAB92574.1; -.
DR InterPro; IPR002559; Transposase_11.
DR Pfam; PF01609; Transposase_11; 1.
FT NON_TER 150 150
SQ SEQUENCE 150 AA; 17286 MW; D38FBC260798DC39 CRC64;

Query Match 11.9%; Score 74; DB 2; Length 150;
Best Local Similarity 29.9%; Pred. No. 2.5;
Matches 29; Conservative 9; Mismatches 33; Indels 26; Gaps 4;

QY 5 PLPVDRAAGLKEKTLILLDYSTKNPVRTVNFNLSQLDPSIHDGWLDFLSSKRLVTLA 64
| : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 4 PNPTDR--GKKXSKIHLLVDRRGLP-----LSIGISAANLHD-----SQALPLV 46
| : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |

QY 65 RGLSPAFLRFGGKRTDFLQFONLRNPAKSRGGPGDPY 101
| : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 47 RGIPPIRSRGRPR------RRPGKLHGDKGIDY 74
| : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |

Search completed: July 30, 2002, 08:30:47
Job time: 999 sec
```


THIS PAGE BLANK (USPTO)


```
;
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Robert Sheinbein
; STREET: 2940 Birchtree lane
; CITY: Silver Spring
; STATE: Maryland
; COUNTRY: United States of America
; ZIP: 20906
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; OPERATING SYSTEM: Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted to
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/922,170B
; FILING DATE: 2 SEP 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedmam, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 910/1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 972-3-5625553
; TELEFAX: 972-3-5625554
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-922-170B-10
;
; Query Match 25.2%; Score 156.5; DB 2; Length 543;
; Best Local Similarity 37.2%; Pred. No. 8.7e-11;
; Matches 42; Conservative 20; Mismatches 44; Indels 7; Gaps 3;
;
; QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFSLQLDPSIIHD-GWLDFLSKR 59
; | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
; Db 22 GPLSPGALPRPA--QAQDVVDLDFFTQEPHLVSPSFLSVTIDANLATDPRFLILGSPK 79
;
; QY 60 LVTLARGLSPAFLRFGKRTDFLOFQNLNPNPAKSRGGPGDPYLYKNVEDDIVR 112
; | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
; Db 80 LRTLARGLSPAYLRFSGTKTDFLIF---DPKKESTFEERSYMQSQVQNQDICK 128
;
; RESULT 3
; US-09-071-739B-2
; Sequence 2, Application US/09071739B
; Patent No. 6177545
; GENERAL INFORMATION:
; APPLICANT: Iris Pecker et al.
; TITLE OF INVENTION: HEPARANASE SPECIFIC MOLECULAR PROBES
; TITLE OF INVENTION: AND THEIR USE IN RESEARCH AND MEDICAL
; TITLE OF INVENTION: APPLICATIONS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
; STREET: 20001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
```

```
;
; OPERATING SYSTEM: Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted to
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,739B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/922,180
; FILING DATE: September 2, 1997
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedmam, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 910/5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 972-3-5625553
; TELEFAX: 972-3-5625554
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-09-071-739B-2
;
; Query Match 25.2%; Score 156.5; DB 4; Length 543;
; Best Local Similarity 37.2%; Pred. No. 8.7e-11;
; Matches 42; Conservative 20; Mismatches 44; Indels 7; Gaps 3;
;
; QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFSLQLDPSIIHD-GWLDFLSKR 59
; | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
; Db 22 GPLSPGALPRPA--QAQDVVDLDFFTQEPHLVSPSFLSVTIDANLATDPRFLILGSPK 79
;
; QY 60 LVTLARGLSPAFLRFGKRTDFLOFQNLNPNPAKSRGGPGDPYLYKNVEDDIVR 112
; | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
; Db 80 LRTLARGLSPAYLRFSGTKTDFLIF---DPKKESTFEERSYMQSQVQNQDICK 128
;
; RESULT 4
; US-09-181-336-13
; Sequence 13, Application US/09181336A
; Patent No. 6242238
; GENERAL INFORMATION:
; APPLICANT: FREEMAN, Craig Geoffrey
; APPLICANT: PARISH, Richard
; APPLICANT: HAMDORF, Brenton James
; APPLICANT: HULETT, Mark Darren
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN
; TITLE OF INVENTION: ENDOGLUCURONIDASE AND USES THEREFOR
; FILE REFERENCE: 032505-004
; CURRENT APPLICATION NUMBER: US/09/181,336A
; CURRENT FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: AU PP0062
; EARLIER FILING DATE: 1997-10-28
; EARLIER APPLICATION NUMBER: AU PP0812
; EARLIER FILING DATE: 1997-12-09
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 13
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Homo sapiens
;
; US-09-181-336-13
;
; Query Match 25.2%; Score 156.5; DB 4; Length 543;
; Best Local Similarity 37.2%; Pred. No. 8.7e-11;
; Matches 42; Conservative 20; Mismatches 44; Indels 7; Gaps 3;
;
; QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFSLQLDPSIIHD-GWLDFLSKR 59
```



```
Db 22 GPLSPGALPRRA--QAQDVLDLFFTOEPLHLVSPSLSVTDANLATDPRFLILGSPK 79
Qy 60 LVTLAGLSPAFLRFGGKRTDFLQFQNLNPAKSRGGPGDYLYLKNYEDD1VR 112
Db 80 LRTLARGLSPAYLRFGGTKTDFLIF---DPKKESTEERSYQSQVYNQDICK 128

RESULT 5
US-09-260-038B-2
; Sequence 2, Application US/09260038B
; Patent No. 6348344
; GENERAL INFORMATION:
; APPLICANT: Maty Ayal-Hershkovitz et al.
; TITLE OF INVENTION: GENETICALLY MODIFIED CELLS AND METHODS FOR
; EXPRESSING RECOMBINANT HEPARANASE
; AND METHODS OF PURIFYING SAME
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
; STREET: 2001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted to
; an ASCII file
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/260,038B
; FILING DATE: 02-Mar-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/071,618
; FILING DATE: May 1, 1998
; APPLICATION NUMBER: 09/071,739
; FILING DATE: May 1, 1998
; APPLICATION NUMBER: 08/922,180
; FILING DATE: September 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedmam, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 910/16
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 972-3-5625553
; TELEFAX: 972-3-5625554
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-260-038B-2

Query Match 25.2%; Score 156.5; DB 4; Length 543;
Best Local Similarity 37.2%; Pred. No. 8.7e-11;
Matches 42; Conservative 20; Mismatches 44; Indels 7; Gaps 3;

Qy 1 GDRRLPVDRAAGLKEKTLILLVSTKNPVRTYNNENFLSLQLDPSIHD-GWLDFLSSKR 59
Db 22 GPLSPGALPRRA--QAQDVLDLFFTOEPLHLVSPSLSVTDANLATDPRFLILGSPK 79
Qy 60 LVTLAGLSPAFLRFGGKRTDFLQFQNLNPAKSRGGPGDYLYLKNYEDD1VR 112
Db 80 LRTLARGLSPAYLRFGGTKTDFLIF---DPKKESTEERSYQSQVYNQDICK 128
```

```
RESULT 6
US-09-102-528-21
; Sequence 21, Application US/09102528
; Patent No. 6207883
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: DNA Sequences
; NUMBER OF SEQUENCES: 32
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/102,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/GB96/03191
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 512 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..512
; OTHER INFORMATION: /note= "Figure 9, sequence of
; AT41a"
US-09-102-528-21

Query Match 10.8%; Score 67; DB 4; Length 512;
Best Local Similarity 27.2%; Pred. No. 6.8;
Matches 25; Conservative 9; Mismatches 24; Indels 34; Gaps 5;

Qy 38 LSLQLDPSI-----IH-----DGLWDLFSSKRLVTLARGLSPA 70
Db 76 LNFQIDPPVQKLYVDIHGAVWDFRHHYRGTPRRHLLTGTGWSKFVNSKKLIA---GDSVW 132
Qy 71 FLREGGKRTDFLQFQNLNPAKSRGGPGDY 102
Db 133 FMR--KSADEMYIGVRRTPISSSDG-GSSYY 160

RESULT 7
US-09-102-528-12
; Sequence 12, Application US/09102528
; Patent No. 6207883
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: DNA Sequences
; NUMBER OF SEQUENCES: 32
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/102,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/GB96/03191
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 584 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```


; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,474
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-9902
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 352 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-751-474-2

Query Match 9.88; Score 61; DB 2; Length 352;
Best Local Similarity 32.3%; Pred. No. 22;
Matches 20; Conservative 8; Mismatches 28; Indels 6; Gaps 3;

QY 32 TVNENFLSLQDPSIHDGW-LDFLSKRLV----TLARGLSPAFLREG-GKRTDFLOFQ 85
Db 11 TVGHVTLNLLMPKFIEDGWEVHYIGDKRGIEHIOELKSLDVTTFHSIATGKLRYPFSWQ 70

QY 86 NL 87
Db 71 NM 72

Search completed: July 30, 2002, 08:15:05
Job time: 287 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:14:00 ; Search time 141.47 Seconds
(without alignments)
387.074 Million cell updates/sec

Title: US-09-836-461-2_COPY_42_534
Perfect score: 2577
Sequence: 1 GDRRLPVDRAAGLKEKTLI.....PVTMGFFVKNYNALACRYR 493

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2577	100.0	534	22 AAB85216	Heparanase-like pr
2	2564	99.5	492	22 AAB84664	Amino acid sequenc
3	2538	98.5	592	22 AAU07424	Human heparanase-1
4	2538	98.5	592	22 AAB85215	Heparanase-like pr
5	2538	98.5	592	22 AAU97632	Human Heparanase,
6	2528	98.1	592	22 AAB81062	Human Heparanase-2
7	2277	88.4	538	22 AAU97633	Human heparanase,
8	2270	88.1	439	22 AAU07423	Human heparanase-1
9	2270	88.1	480	22 AAU07418	Novel human extrac
10	2270	88.1	480	22 AAB85217	Heparanase-like pr
11	2270	88.1	480	22 AAU97634	Human heparanase,

12	1112	43.2	214	22	AAM99905	Human excretory re
13	1112	43.2	214	22	AAAM3704	Human bladder anti
14	1106	42.9	262	22	AAAM2417	Human EST encoded
15	944.5	36.7	535	21	AAB08851	A murine heparanas
16	930.5	36.1	543	20	AAU02345	A human heparanase
17	930.5	36.1	543	21	AAB08849	Amino acid sequenc
18	930.5	36.1	543	21	AAU57590	Human heparanase.
19	930.5	36.1	543	21	AAU52990	Human heparanase p
20	930.5	36.1	543	22	AAU97635	Human heparanase p
21	930.5	36.1	543	22	AAU02346	A human heparanase
22	930.5	36.1	592	21	AAB08850	Amino acid sequenc
23	927.5	36.0	543	22	AAB88361	Human membrane or
24	926.5	36.0	530	20	AAU34173	Human pre-prohepar
25	926.5	36.0	543	20	AAU17082	Human heparanase e
26	926.5	36.0	543	22	AAB86206	Human heparanase i
27	926.5	36.0	588	20	AAU30124	Human heparanase i
28	883.5	34.3	532	20	AAU17083	A human protein wi
29	805	31.2	380	20	AAU17085	Seq ID No: 15 of W
30	779	30.2	380	20	AAU17084	Rat heparanase enz
31	778	30.2	156	22	AAG65963	Mouse heparanase e
32	468.5	18.2	488	22	AAB31459	Human heparanase-1
33	440.5	17.1	488	22	AAB31472	Amino acid sequenc
34	439.5	17.1	488	22	AAB31470	Amino acid sequenc
35	422.5	16.4	488	22	AAB31471	Amino acid sequenc
36	278	10.8	137	22	AAG65961	Human heparanase-1
37	278	10.8	159	22	AAG65964	Human heparanase-1
38	270	10.5	112	22	AAU07425	Human heparanase-1
39	247	9.6	256	21	AAG13479	Arabidopsis thalia
40	203	7.9	118	22	AAG65962	Human heparanase-1
41	171	6.6	935	22	AB869219	Drosophila melanog
42	162	6.3	115	22	AAB85218	Mouse heparanase-1
43	144	5.6	38	20	AAU34186	Human pre-prohepar
44	127	4.9	24	22	AAB85220	Heparanase-like pr
45	127	4.9	25	22	AAB85221	Heparanase-like pr

ALIGNMENTS

RESULT	1
AAB85216	AAAB85216 standard; Protein; 534 AA.
ID	AC AAB85216;
AC	07-SEP-2001 (first entry)
DT	Heparanase-like protein Hpa2 splice variant #2.
DE	Heparanase; splice variant; homologue; heparanase-like protein; Hpa2; cytosolic; neuroprotective; cerebroprotective; immunosuppressive; antipsoriatic; nontropic; antiinflammatory; antiarthritic; antiasthmatic; antidiabetic; antiarteriosclerotic; vulnery.
KW	Homo sapiens.
KW	WO200146392-A2.
PN	28-JUN-2001.
PD	21-DEC-2000; 2000WO-GB04963.
PF	22-DEC-1999; 99GB-0030392.
PR	07-APR-2000; 2000GB-0008713.
XX	(OXFO-) OXFORD GLYCOSCIENCES UK LTD.
PA	McKenzie EA, Stamps AC, Terrett JA, Tyson KL;
XX	WPI; 2001-418056/44.
DR	N-PSDB; AAH22672.
XX	Novel homologs of heparanase, present in three splice variants, useful
PT	

PT for identifying agents that modulate heparanase, useful in the
PT treatment and/or prophylaxis of abnormal levels of heparanase
XX
XX
XX Claim 1; Fig 2; 97pp; English.
XX
XX The invention provides a homologue to heparanase which is present in
CC three splice variants. The heparanase homologue polypeptide is useful in
CC the treatment of a human or non-human animal or for use in diagnosis.
CC Vectors comprising the heparanase homologue polynucleotides are useful in
CC the transformation or transfection of a prokaryotic or eukaryotic host.
CC The modulators of the polypeptide are useful in the manufacture of a
CC medicament for the treatment and/or prophylaxis of a condition/disease
CC associated with abnormal levels of the heparanase homologue, including
CC cancer, central nervous system (CNS) and neurodegenerative diseases.
CC cardiovascular diseases such as restenosis following angioplasty and
CC atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus,
CC allografts, inflammatory diseases, arthritis, vascular restenosis,
CC tumour growth and progression, asthma, Alzheimer's disease, diabetic
CC retinopathy, wound healing and inflammation. The polypeptide is also
CC useful in diagnosis and research. The present sequence represents the
CC amino acid sequence of the mid-sized splice variant of the heparanase-
XX like protein Hpa2 of the invention.
XX
SQ Sequence 534 AA;

Query Match 100.0%; Score 2577; DB 22; Length 534;
Best Local Similarity 100.0%; Pred. No. 3.8e-252;
Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDRRLPVDRAGLKEKTLILLDVSTKNPVRTVNENFLSLQDPSIIHDGWLDFLSSKRL 60
DB 42 GDRRLPVDRAGLKEKTLILLDVSTKNPVRTVNENFLSLQDPSIIHDGWLDFLSSKRL 101

QY 61 VTLAGLSPAFIRFGGKRTDFLQFONLRNPAKSRGGPGDPYILKNYEDDVRSDVALDKQ 120
DB 102 VTLAGLSPAFIRFGGKRTDFLQFONLRNPAKSRGGPGDPYILKNYEDDVRSDVALDKQ 161

QY 121 GKCKTAQHPDVMLEQREKAQMHLVLLKEQFSNTYSNLILTEPNYRTMHGRAVNGSQL 180
DB 162 GKCKTAQHPDVMLEQREKAQMHLVLLKEQFSNTYSNLILTEPNYRTMHGRAVNGSQL 221

QY 181 GKDYLQKSLQPIRIYSRASLYGPNIGRPRKNVIALDGFMKVAGSVDAVTHQHCYID 240
DB 222 GKDYLQKSLQPIRIYSRASLYGPNIGRPRKNVIALDGFMKVAGSVDAVTHQHCYID 281

QY 241 GRVVKVMDFLKRLDITLSDQIRKIQKVNTYTPGKKIWLCEVYVTSAGGTNNLSDSYAA 300
DB 282 GRVVKVMDFLKRLDITLSDQIRKIQKVNTYTPGKKIWLCEVYVTSAGGTNNLSDSYAA 341

QY 301 GFLWLTNTLQMLANQGVVIRHSFDDHGVNHLVDQFNPLPDYWLSLLYKRLIGPKVLAV 360
DB 342 GFLWLTNTLQMLANQGVVIRHSFDDHGVNHLVDQFNPLPDYWLSLLYKRLIGPKVLAV 401

QY 361 HVAGLQKRPGRVIRDKLRIYAHCTNNHNNYVRGSTTLFINLHRSRKKTKLAGTLRD 420
DB 402 HVAGLQKRPGRVIRDKLRIYAHCTNNHNNYVRGSTTLFINLHRSRKKTKLAGTLRD 461

QY 421 KLHVQYLLOPYQEGELKSKVOLNGOPLVMVDGTLPELKPRLRAGRTLVIPTVTMGFF 480
DB 462 KLHVQYLLOPYQEGELKSKVOLNGOPLVMVDGTLPELKPRLRAGRTLVIPTVTMGFF 521

QY 481 VKNVNALACRYR 493
DB 522 VKNVNALACRYR 534

RESULT 2
ID AAB84664
XX AAB84664 standard; Protein; 492 AA.
AC AAB84664;
XX

DT 05-SEP-2001 (first entry)
DE Amino acid sequence of human heparanase-like polypeptide.
XX
XX Human; heparanase-like polypeptide; gene therapy; cancer; angiogenesis;
KW trauma; autoimmune disease; skin disease; cardiovascular disease;
KW nervous system disease; inflammation; arthritis; genitalia;
KW male fertility; erectile dysfunction.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Misc-difference 407 /note= "unspecified residue encoded by KCA"
FT
FT
XX WQ200148161-A2.
XX
PD 05-JUL-2001.
XX
PF 18-DEC-2000; 2000WO-EP12909.
XX
PR 23-DEC-1999; 99EP-0125831.
XX
XX (SCHD) SCHERING AG.
XX
XX Siemeister G, Weiss B;
PI
XX WPI: 2001-418259/44.
DR N-PSDB: AAH28347.
DR
XX Human Heparanase-like polynucleotide encoding polypeptides useful for
PT modulating expression of the polypeptide and for treating cancer, -
PT cancer metastasis, aberrant angiogenesis by gene therapy technique -
XX
XX Claim 9; Page 30; 30pp; English.
PS
XX The present sequence represents a human heparanase-like polypeptide.
CC Heparanase-like polynucleotides are useful as a source of probes,
CC primers and antisense molecules, and in gene therapy. Heparanase-like
CC polynucleotides and polypeptides are useful for treating several
CC disorders e.g., cancer, cancer metastasis. The oligonucleotides are
CC also useful as diagnostic markers for the diagnosis of disorder such
CC as cancer, cancer metastasis and aberrant angiogenesis. They may also
CC act as diagnostic markers for diagnosis of disorder such as cancer,
CC cancer metastasis and aberrant angiogenesis. The heparanase polypeptides
CC and polynucleotides are also useful for treating trauma, autoimmune
CC diseases, skin diseases, cardiovascular diseases, nervous system
CC diseases, and inflammation including arthritis. Since the polynucleotide
CC is preferentially expressed in male genitalia, modulation of its
CC expression and/or activity may be used for medical intervention in male
CC genitalia function that is male fertility control, erectile dysfunction.
XX
XX Sequence 492 AA;

Query Match 99.5%; Score 2564; DB 22; Length 492;
Best Local Similarity 99.6%; Pred. No. 6.9e-251;
Matches 490; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 DRRPLPVDRAGLKEKTLILLDVSTKNPVRTVNENFLSLQDPSIIHDGWLDFLSSKRLV 61
DB 1 DRRPLPVDRAGLKEKTLILLDVSTKNPVRTVNENFLSLQDPSIIHDGWLDFLSSKRLV 60

QY 62 TLARGLSPAFIRFGGKRTDFLQFONLRNPAKSRGGPGDPYILKNYEDDVRSDVALDKQK 121
DB 61 TLARGLSPAFIRFGGKRTDFLQFONLRNPAKSRGGPGDPYILKNYEDDVRSDVALDKQK 120

QY 122 GKCKTAQHPDVMLEQREKAQMHLVLLKEQFSNTYSNLILTEPNYRTMHGRAVNGSQL 181
DB 121 GKCKTAQHPDVMLEQREKAQMHLVLLKEQFSNTYSNLILTEPNYRTMHGRAVNGSQL 180

QY 182 KDYLQKSLQPIRIYSRASLYGPNIGRPRKNVIALDGFMKVAGSVDAVTHQHCYIDG 241
DB 182 KDYLQKSLQPIRIYSRASLYGPNIGRPRKNVIALDGFMKVAGSVDAVTHQHCYIDG 241

Db 181 kdiqlkllqpiriyraslygnprprknvialldgfmkvagstvdavtwqhcyyidg 240
Qy 242 RVKVMDFKRLDLTSDQIRKIQKVNTYTPGKKIWLEGVVTTSAGGTNNLSDSAAG 301
Db 241 rrvkvmdfkrlldtisdqirkikqvntytpgkklwlegvvttsaggtnnlsdyaag 300
Qy 302 FLWLNTGLMANGIDVIRSFDFHGYNHLVDQFNPLPDYWLISLLYKRLIGPKVLAVH 361
Db 301 flwntlgmlanggidvvrhsfddghgnhlvdqfnplpdywlslllykrligpkvlavh 360
Qy 362 VAGLQRPGRVIRDKLRIYAHCTNHHNHNKVRGSIPLFIINLHRSRKKIKLAGTLRDK 421
Db 361 vaglqrprgrvirdklriyahctnhhnhvrgsiltfiinlhrxrkklkgtlrdk 420
Qy 422 LVHQLLPYGOEGLKSKSVQLNGOPLVMVDGTLPELKPRLRAGRTLVIPTVTMGFFV 481
Db 421 lvhqllypygqeglkksvqlngqplvmvddgtlpeikprlragrtlviptvtmgffv 480
Qy 482 VKNVNALACRYR 493
Db 481 vknvnalacryr 492
RESULT 3
AAU07424
ID AAU07424 standard; Protein: 592 AA.
XX
AC
XX
XX
DT 18-DEC-2001 (first entry)
XX
DE Human heparanase-like protein splice variant #1.
XX
KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW antiproliferative; cardiac; vasotropic; cerebroprotective; neurotropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; cardiovascular disease;
KW nervous system disorder; Alzheimer's disease; cerebral ischaemia; infection;
KW wound healing; food additive; heparanase.
XX
OS Homo sapiens.
XX
XX WO200179253-A1.
XX
XX
PD 25-OCT-2001.
XX
XX 11-APR-2001; 2001WO-US11643.
PF
XX
PR 18-APR-2000; 2000US-198123P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Fiscella M, Shi Y, Ebner R, Ruben SM;
XX
DR WPI; 2001-611720/70.
DR N-PSDB; AAS13848.
XX
PT New nucleic acids encoding extracellular matrix polypeptides, for
PT diagnosing, treating, preventing or ameliorating human disorders and
PT disease, such as, autoimmune, hyperproliferative or cardiovascular
PT disorders -
XX
PS Disclosure; Page 14; 308pp; English.
XX
CC The invention relates to novel isolated polynucleotides (I) encoding
CC extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by
CC (I) are used to prevent, treat or ameliorate a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility to
CC a pathological condition. The antibodies to the polypeptides can also be
CC used in alleviating symptoms associated with the disorders and in

CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease. Infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities. The present sequence represents the amino acid sequence
CC of human heparanase-like protein, splice variant #1.
XX
SQ Sequence 592 AA;
Query Match 98.5%; Score 2538; DB 22; Length 592;
Best Local Similarity 89.5%; Pred. No. 4e-248;
Matches 493; Conservative 0; Mismatches 0; Indels 58; Gaps 1;
Qy 1 GDRRLPLVDRAAGLKEKTLILLDVSTKPNPRTVNENFSLQLDPSIIHDGWLDFLSSKRL 60
Db 42 gdrpripvdraaglkettillldvstknprtvtvnfnslqldpsilhdgwlidflsskrl 101
Qy 61 VTLARGLSPAFLRFGCKRTDFLQFONLRNPAKSRGCPGPDYVYKKNVEDDIVRSDVALDKQ 120
Db 102 vtlarglspaflrfgkrtdfllqfqlnrnpaksr99pdyvlykneyddivrsvdvaldkq 161
Qy 121 KGCKTAQHPDVMLELQREKAAQMHLVLKKEQFSNTYSNLILT----- 162
Db 162 kgckiaqhpdvmllelqrekaqmhlvlkkgfsntysnllltcarsldklynfadcsglhl 221
Qy 163 -----EPNNYRTMHGRAVNSQLGK 182
Db 222 ifalnrlrnpnnswnssalsllkysaskkyniswelgnepnnyrtmhgravngsqlgk 281
Qy 183 DYIQLKSLLOPTIRYSRASLYGNIGRPNKVNIALLDGFMKVAGSTVDVAVTWHQHCYIDGR 242
Db 282 dyiqklsllqpiriyraslygnprprknvialldgfmkvagstvdavtwqhcyyidgr 341
Qy 243 VVKVMDFLKRLDLTSDQIRKIQKVNTYTPGKKIWLEGVVTTSAGGTNNLSDSAAGF 302
Db 342 vvkvmdfkrlldtisdqirkikqvntytpgkklwlegvvttsaggtnnlsdyaagf 401
Qy 303 LWLNTGLMANGIDVIRSFDFHGYNHLVDQFNPLPDYWLISLLYKRLIGPKVLAVHV 362
Db 402 lwlntlgmlanggidvvrhsfddghgnhlvdqfnplpdywlslllykrligpkvlavhv 461
Qy 363 AGLQRPGRVIRDKLRIYAHCTNHHNHNKVRGSIPLFIINLHRSRKKIKLAGTLRDKL 422
Db 462 aglqrprgrvirdklriyahctnhhnhvrgsiltfiinlhrxrkklkgtlrdk 521
Qy 423 VHQLLPYGOEGLKSKSVQLNGOPLVMVDGTLPELKPRLRAGRTLVIPTVTMGFFV 482
Db 522 vhqyllpygqeglkksvqlngqplvmvddgtlpeikprlragrtlviptvtmgffv 581
Qy 483 KNVNVALACRYR 493
Db 582 knvnalacryr 592
RESULT 4
AAB85215
ID AAB85215 standard; Protein: 592 AA.
XX
AC AAB85215;
XX
DT 07-SEP-2001 (first entry)
XX
DE Heparanase-like protein Hpa2 splice variant #1.

```
xx Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;
kw cytosolic; neuroprotective; cerebroprotective; immunosuppressive;
kw antiproliferative; neurotropic; antiinflammatory; antiarthritic; antiasthmatic;
kw antidiabetic; antiarteriosclerotic; vulnerary.
xx
OS Homo sapiens.
FH
FT Key Location/Qualifiers
FT Misc-difference 237
FT /label= unknown
FT /note= "encoded by ANC"
xx
xx WO200146392-A2.
xx
xx 28-JUN-2001.
xx
xx 21-DEC-2000; 2000WO-GB04963.
xx
xx 22-DEC-1999; 99GB-0030392.
xx 07-APR-2000; 2000GB-0008713.
xx
xx (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
xx
xx McKenzie EA, Stamps AC, Terrett JA, Tyson KL;
xx WPI; 2001-418056/44.
xx N-PSDB; AAH22671.
xx
xx Novel homologs of heparanase, present in three splice variants, useful
xx for identifying agents that modulate heparanase, useful in the
xx treatment and/or prophylaxis of abnormal levels of heparanase
xx
xx Claim 1; Fig 1; 97pp; English.
xx
xx The invention provides a homologue to heparanase which is present in
xx three splice variants. The heparanase homologue polypeptide is useful in
xx the treatment of a human or non-human animal or for use in diagnosis.
xx Vectors comprising the heparanase homologue polynucleotides are useful in
xx the transformation or transfection of a prokaryotic or eukaryotic host.
xx The modulators of the polypeptide are useful in the manufacture of a
xx medicament for the treatment and/or prophylaxis of a condition/disease
xx associated with abnormal levels of the heparanase homologue, including
xx cancer, central nervous system (CNS) and neurodegenerative diseases,
xx cardiovascular diseases such as restenosis following angioplasty and
xx allografts, autoimmune diseases, psoriasis, lupus erythematosus,
xx atherosclerosis, inflammatory diseases, arthritis, vascular restenosis,
xx tumour growth and progression, asthma, Alzheimer's disease, diabetic
xx retinopathy, wound healing and inflammation. The polypeptide is also
xx useful in diagnosis and research. The present sequence represents the
xx amino acid sequence of the largest splice variant of the heparanase-like
xx protein Hpa2 of the invention.
xx
xx Sequence 592 AA;
xx
xx Query Match 98.5%; Score 2538; DB 22; Length 592;
xx Best Local Similarity 89.5%; Pred. No. 4e-248;
xx Matches 493; Conservative 0; Mismatches 0; Indels 58; Gaps 1;
xx
xx 1 GDRRLPVDRAGLUKETLLILDVSTKNPVRTVNEFLSLQLOPSIHDWLDLSKRL 60
xx
xx 42 gdrplpvdraaglketllildvstknpvrtnenflslqldpsiihdgldflskrl 101
xx
xx 61 VTLARGLSPALRFGGRTDFLOFNLRNPAKSRGGPGDYILKNYEDDILVRSDVALDKQ 120
xx
xx 102 vtlarglsapalrfggrrtdflqfnlnpaksrsggpgdyilknycddivrsdvaldkq 161
xx
xx 121 KGCKIAQHPDVNLELQREKAQMHLVLLKQFSNTYSNLIILT----- 162
xx
xx 162 kgckiaqhpdvmlqrekaqmhlvllkqfsntysnliiltarsldklynfadcsghl 221
xx
xx 163 -----EPNNRYMTMHGRAVNSQLGK 182
```

```
Db 222 ifalnairnnpnnsxssalsllkysaskkyniswelgnepnnnyrtmhgravngslgk 281
Qy 183 DYIOLSKLQPIRIYGRASYGNIGRPRKNVTALLDGFMKVGASTVDATWQHCVIDGR 242
Db 282 dyiqklsllqpiriyrasygnigrprknvialldgfmkvagstvdatwqncylodr 341
Qy 243 VVKVDFLTKRLDLSQIRKIQKVVNTYTPGKKIWLQGVVTTSGAGTNNLSDSYAAGF 302
Db 342 vvkvmdfiktrlltdlsdqirkkqvntvtycpkklwlegvvttsaggttnladsyaagf 401
Qy 303 LWNLTGLMLANQGDVVIRHSFFDHGYNHLVDQNFNLPDYWLSLLYKRLIGPKVLAVHV 362
Db 402 lwntlgmlanqgdvvrhsffdhgynhlvdqnfnpdywlslllykrligpkvlavhv 461
Qy 363 AGLQKRPGRVIRDLKRIYAHCTNHHNHNHYVRSITFLFIINLHRSRKKIKLAGTIRDKL 422
Db 462 aglqrkprgrvirdkrlriyahctnhhnhyvrsitflfiinhrrskkkiklagtirdkl 521
Qy 423 VHQYLLQPYGQGLSKSVQNLNGQPLVMVDGTLPELKPRLRAGRTLVIPTVMTGFFVV 482
Db 522 vhyllqpygqglksksvqlngqplvmvdgtlpeikprlragrtlvipvtmgffvv 581
Qy 483 KVNVALACRYR 493
Db 582 knvnalacryr 592
xx
xx RESULT 5
xx AAY97632
xx ID AAY97632 standard; Protein; 592 AA.
xx AC AAY97632;
xx XX
xx DT 20-APR-2001 (first entry)
xx DE Human heparanase, hnhp1, protein sequence.
xx KW Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;
xx KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
xx KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
xx KW gene therapy; human.
xx OS Homo sapiens.
xx PN WO200100643-A2.
xx PD 04-JAN-2001.
xx XX
xx PF 19-JUN-2000; 2000WO-IL00358.
xx XX
xx PR 25-JUN-1999; 99US-0140801.
xx XX
xx PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
xx PI Pecker I, Michal I, Itzhaki H;
xx XX
xx DR WPI; 2001-137930/14.
xx DR N-PSDB; AAA91097.
xx XX
xx PT New polynucleotides and polypeptides that are distantly homologous to
xx PT heparanase, useful in wound healing, as well as in gene therapy
xx PT protocols for angiogenesis, restenosis, atherosclerosis, or
xx PT inflammation -
xx XX
xx PS Claim 10; Fig 1; 67pp; English.
xx PS
xx CC This sequence represents a heparanase of the invention.
xx CC The heparanase DNA and protein sequences are useful in wound healing,
xx CC angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary
xx CC diseases, neurodegenerative diseases (such as Scrape, Alzheimer's
xx CC disease, and Creutzfeldt-Jakob disease) or viral infections. The
xx CC heparanase coding sequence is particularly useful in gene therapy.
```

XX	SQ	Sequence	592 AA;	
Query Match				
Best Local Similarity 98.5%; Score 2538; DB 22; Length 592;				
Matches 493; Conservative 0; Mismatches 0; Indels 58; Gaps 1;				
QY	1	GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNEFLSLQLDPSIIHDGWLDFLSKRL	60	
Db	42	gdrpvpdraaglkectllildvstknprtvenfnlsldqpsihdgwldflsskrl	101	
QY	61	VTLAGLSPAFIRFGGKRTDFLQFNLRNPAKSRGGPGPDYILKNYEDDIVRSDVALDKQ	120	
Db	102	vtlarglspafirfggkrtdfiqfnlrnpaksrpgpdyilknyeddivrdsvaldkq	161	
QY	121	KGCKIAQHPDVMLELQREAAQMHLVLLKEQFSNTYSNLILT-----	162	
Db	162	kgckiaqhpdvmllelqrekaaqmhlvllkeqfsntysnliltarsldklynfadcsghl	221	
QY	163	-----EPNNYRTMHGRAVNGSQLGK	182	
Db	222	ifalnallrnpnnswnssalsllkysaskkyniswelgnepnnyrtmhgravnsgqlgk	281	
QY	183	DYIQLKSLLOPIRIYSRASLYGNIPRKNVIALLDGFMKVAGSTVDVATWQHCVIDGR	242	
Db	282	dyiqlksllqpiriysraslygniprknvialldgfmkvagstvdavtwhqcyidgr	341	
QY	243	VVKVMDFLKTRLLDTSQIRKIQKVNTYTPCKKIWLKGVVTTTSAGGTTNLSDSYAAGF	302	
Db	342	vvkvmdfktrlldtsdqirkiqkvntytpgkklwlegvvtttsaggttnlsdsyaagf	401	
QY	303	LWLNTGLMLANGIDVIRHSPFDHGYNHLVDQNFNPLPDYILSLYKRLIGPKVLAVHV	362	
Db	402	lwlntglmlangidvirhsffdhgynhlvdqfnplpdyilsllykrligpkvlavhv	461	
QY	363	AGLQKRPGRVIRDKLRIYACHTNNHNNYVSGSTLFIINLHRSRKKIKLAGTLURDKL	422	
Db	462	aglqkrpgrvirdklriyachtnhnnhnyvrgstlfiinlhrsrrkkiklagtlurdkl	521	
QY	423	VHQLLPQYQEGLSKSVOLNGQPLVMYDDGTLPELKPRPRAGRTLVIPTVMGFFVY	482	
Db	522	vhqyllpqyqeglsksvqlngqplvmvddgtlpelkprpragrtlviptvmgffv	581	
QY	483	KNVNALACRYR 493		
Db	582	knvnalacryr 592		
RESULT 6				
ID	AAB81062 standard; Protein; 592 AA.			
XX	AAB81062;			
DT	20-JUN-2001 (first entry)			
XX	Human Heparanase-2 amino acid sequence.			
DE	Heparanase 2: human; endoglucuronidase; heparan sulphate; metastasis;			
KW	neangiogenesis; vaccine; autoimmune disorder; blood coagulation;			
KW	cancer; diabetes; ischaemia; sepsis; stroke; cardiovascular; thrombosis.			
XX	Homo sapiens.			
OS				
Key	Location/Qualifiers			
PH	156..169			
FT	/label= Immunogenenic_epitope			
FT	249..262			
FT	/label= Immunogenenic_epitope			
FT	505..518			
FT	/label= Immunogenenic_epitope			
XX				

PN	WO200121814-A1.			
XX				
PD	29-MAR-2001.			
XX				
PF	11-SEP-2000; 2000WO-EP08837.			
XX				
PR	23-SEP-1999; 99EP-0118805.			
PR	07-JUL-2000; 2000EP-0114649.			
XX	(MERE) MERCK PATENT GMBH.			
PA				
PI	Duecker K, Sirrenberg C;			
XX				
DR	WPI; 2001-308089/32.			
DR	N-PSDB; AAP86101.			
XX				
PT	New heparanase-2 polypeptide useful in diagnosing (the susceptibility			
PT	of a subject to) and as vaccines against e.g. autoimmune disorders,			
PT	cardiovascular disease, cancer, diabetes, ischemia, sepsis, stroke, or			
PT	thrombosis			
XX				
PS	Claim 1; Page 42-43; 46pp; English.			
XX				
CC	This invention relates to a human heparanase-2 protein and the cDNA			
CC	sequence encoding it. Heparanase-2 is a member of the endoglucuronidase			
CC	family of polypeptides and it degrades heparan sulphate proteoglycans			
CC	HSPGs (ubiquitous macromolecules of cell surfaces, basement membranes and			
CC	the extracellular matrix). HSPGs support the vascular endothelium and			
CC	stabilise the structure of the capillary wall. Heparanases may be			
CC	associated with neangiogenesis and metastasis related to malignant			
CC	tumour formation. Heparanase-2 polynucleotides and proteins are useful as			
CC	vaccines for inducing an immunological response against autoimmune			
CC	disorders, blood coagulation disorders, cancer, diabetes, ischaemia,			
CC	sepsis, stroke, cardiovascular diseases, or thrombosis, as well as in			
CC	diagnosing (the susceptibility of a subject to) these diseases.			
CC	Heparanase-2 fragments may be used as immunogens to produce antibodies			
CC	immunospesific to the polypeptides, and to identify membrane bound			
CC	soluble receptors, agonists or antagonists that compete with the binding			
CC	of the polypeptide to the receptors. An antibody specific for			
CC	heparanase-2 can be used in the diagnosis of the above diseases and in			
CC	isolating or identifying clones expressing heparanase-2. The present			
CC	sequence represents heparanase-2. Three regions of heparanase-2 with high			
CC	immunogenicity (immunogenic epitopes) can be used to raise antibodies			
CC	against heparanase-2.			
XX				
SQ	Sequence 592 AA;			
Query Match 98.1%; Score 2528; DB 22; Length 592;				
Best Local Similarity 89.1%; Pred. No. 4.2e-247;				
Matches 491; Conservative 1; Mismatches 1; Indels 58; Gaps 1;				
QY	1	GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNEFLSLQLDPSIIHDGWLDFLSKRL	60	
Db	42	gdrpvpdraaglkectllildvstknprtvenfnlsldqpsihdgwldflsskrl	101	
QY	61	VTLAGLSPAFIRFGGKRTDFLQFNLRNPAKSRGGPGPDYILKNYEDDIVRSDVALDKQ	120	
Db	102	vtlarglspafirfggkrtdfiqfnlrnpaksrpgpdyilknyeddivrdsvaldkq	161	
QY	121	KGCKIAQHPDVMLELQREAAQMHLVLLKEQFSNTYSNLILT-----	162	
Db	162	kgckiaqhpdvmlvlqrekaaqmhlvllkeqfsntysnliltarsldklynfadcsghl	221	
QY	163	-----EPNNYRTMHGRAVNGSQLGK	182	
Db	222	ifalnallrnpnnswnssalsllkysaskkyniswelgnepnnyrtmhgravnsgqlgk	281	
QY	183	DYIQLKSLLOPIRIYSRASLYGNIPRKNVIALLDGFMKVAGSTVDVATWQHCVIDGR	242	
Db	282	dyiqlksllqpiriysraslygniprknvialldgfmkvagstvdavtwhqcyidgr	341	
QY	243	VVKVMDFLKTRLLDTSQIRKIQKVNTYTPGKKIWLKGVVTTTSAGGTTNLSDSYAAGF	302	

XX DR WPI; 2001-611720/70.
XX
PT New nucleic acids encoding extracellular matrix polypeptides, for
PT diagnosing, treating, preventing or ameliorating human disorders and
PT disease, such as, autoimmune, hyperproliferative or cardiovascular
PT disorders -
XX
PS Disclosure; Page 13-14; 308pp; English.
XX
XX The invention relates to novel isolated polynucleotides (I) encoding
CC extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by
CC (I) are used to prevent, treat or ameliorate a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility to
CC a pathological condition. The antibodies to the polypeptides can also be
CC used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC transplant, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities. The present sequence represents the amino acid sequence
CC of human heparanase-like protein.
XX
SQ Sequence 439 AA;

Query Match 88.13; Score 2270; DB 22; Length 439;
Best Local Similarity 89.08; Pred. No. 3.8e-221;
Matches 439; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

QY 1 GDRPLPVDRAAGLKETLILLDYSTKNPVTNENFLSLQDPSIHDGWLDFLSSKRL 60
Db 1 gdrplpvdraaglkettillldydstknpvtntnenflslqdpstilhdgwlflsskrl 60

QY 61 VTLARGSPAFRFGKRTDFLOFNLRNPAKSRGGPGDYIYLNKYEDDVRSDVALDKQ 120
Db 61 vtlargspafirfggkrtdflofnlrnpaksrpggpgdyiynkyed----- 108

QY 121 KGCKIAQHPDVMLEQREKAAQMHLLVLLKEQFSNTYSNLITPENNYRTMHGRAVNGSQL 180
Db 109 -----epnnnyrtmhgravnsgql 126

QY 181 GKDYIQLKSLLOPIRIYSRASLYGNTPGRKNVIALLDGFMKVGASTVDATVWQHCVIID 240
Db 127 gkdyiqklsllqpiriysraslygnpgrknvialldgfmkvagstvdatvwhqcviid 186

QY 241 GRVVKVMDFLKRLDLSQIRKIQKVVNTYTPGKKIWLGVVTTTSAGGTNNLSDSYAA 300
Db 187 grvvkvmfdlkrlltdlsdqirkiqkvvntytpgkklwlgvvtttsaggtnnlstdsyaa 246

QY 301 GFLMNLTLGLMANGIDVIRHSFDFGYNHLVDQNFNPLDPYWLSLLYKRLIGPKVLAV 360
Db 247 gflwntlglmangidvirsfdhgnhlvdqnfnpdpwywlslykrligpkvlav 306

QY 361 HVAGLQRPGRVIRDLRIYAHTNHNHNYVRGSTTLFLINLHRSKKIKLAGTLRD 420
Db 307 hvaglrpgrvirdlriyahctnhnnyvrgsttflflnlhrskkiklagtlrd 366

QY 421 KLVHQYLLQPYQGBGLSKSVQLMGQPLVMVDGTLPELKPRLRAGRRLVIPPVTMGFF 480
Db 367 klvhqyllqpyqgbglsksvqlmgqplvmvddgtlpeklprrlragrtrlvippvtmgff 426

QY 481 VVKNNVALACRYR 493

Db 427 vvknnvalacryr 439
|||||
RESULT 9
AAU07418
ID AAU07418 standard; Protein; 480 AA.
XX
AC AAU07418;
XX
DT 18-DEC-2001 (first entry)
DE DE
XX Novel human extracellular matrix (ECM) protein #1.
KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW antiproliferative; cardiac; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;
KW nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;
KW wound healing; food additive.
XX
OS Homo sapiens.
XX WO200179253-A1.
PN
PD 25-OCT-2001.
XX
PF 11-APR-2001; 2001WO-US11643.
XX
PR 18-APR-2000; 2000US-198123P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Fiscella M, Shi Y, Ebner R, Ruben SM;
XX
DR WPI; 2001-611720/70.
DR N-PSDB; AAS13843.
XX
PT New nucleic acids encoding extracellular matrix polypeptides, for
PT diagnosing, treating, preventing or ameliorating human disorders and
PT disease, such as, autoimmune, hyperproliferative or cardiovascular
PT disorders -
XX
PS Claim 1; Page 292-293; 308pp; English.
XX
CC The invention relates to novel isolated polynucleotides (I) encoding
CC extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by
CC (I) are used to prevent, treat or ameliorate a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility to
CC a pathological condition. The antibodies to the polypeptides can also be
CC used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities. The present sequence represents the amino acid sequence
CC of novel human extracellular matrix (ECM) protein #1.
XX
SQ Sequence 480 AA;

Query Match	88.1%;	Score 2270;	DB 22;	Length 480;
Best Local Similarity	89.0%;	Pred. No. 4.4e-221;		
Matches 439;	Conservative 0;	Mismatches 0;	Indels 54;	Gaps 1;
QY	1	GDRRLPVDRAAGLKEKTLILLDVSTKNPVTYNENFLSLQLDPSIIHDGWLDFLSKRL	60	
DB	42	gdrplpvdraaglikektilldvstknprtvenflslqldpsiihdgwlflskrl	101	
QY	61	VTLAGRLSPAFIRFGGKRTDFLQFNLRNPAKSRGGPGPDYILKNYEDDIVRSDVALDKQ	120	
DB	102	vtlarglspafirfggkrtdflqfnlrnpaksrpgpdyilknyed-----	149	
QY	121	KGCKIAQHPDVMLELQREKAAQMHVLVLLKEQFSNTYSNLILTEPNNTYRTHMGRVNGSQL	180	
DB	150	-----epnnyrtmhgravnsgql	167	
QY	181	KDXYIQLKSLLOPIRIYRASLYGNIGRPKNVIALLDGFMKVAGSTVDVATVWQHCVYID	240	
DB	168	gkdyiqkllqpiriyraslygnigrprknvialldgfmkvagstvdatvqhcyid	227	
QY	241	GRVVKVMDFLKTRLLDLSDDQIRKIQKVVNTYTPGKKIWLGVVTTTSAGGTNNLSDSYAA	300	
DB	228	grvkvmdflktrlldtsdqirkiqkvntytgpkkiwlegvvtttsaggtnnlsdsyaa	287	
QY	301	GFLWLTGLMLANQGDVYIRHSFDDHGYNHLVDQNFNPLDPDYLLYLKRLIGPKVLAV	360	
DB	288	gflwltglmlanqgdvvyirhsfddhgynhlvdgnfnpldpdyllslylkrllgpkvlav	347	
QY	361	HVAGLQRPGRVIRDKLRIYAHCTNHNHNHYVSGTTLFTINLHRSRKKIKLAGTLRD	420	
DB	348	hvaglrkprgrvirdklriyahctnhhnhnyvrgsitlfiinlhrsrrkiklagtlrd	407	
QY	421	KLHVQYLLQPYGQGLKSKSVOLNGOPLVMVDDGTLPDLKPRPLRAGRTLVIPTVMTGFF	480	
DB	408	klvhqyllqpyggeglksksvqlnggplvmvddgtlplkprplragrtlvipvptmgff	467	
QY	481	VKKNVNALACRYR	493	
DB	468	vvkvnvalacryr	480	
RESULT 10				
AAB85217				
ID	AAB85217 standard; Protein; 480 AA.			
XX				
AC	AAB85217;			
XX				
DT	07-SEP-2001 (first entry)			
XX				
DE	Heparanase-like protein Hpa2 splice variant #3.			
XX				
KW	Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;			
KW	cytostatic; neuroprotective; cerebroprotective; immunosuppressive;			
KW	antipsoriatic; neutropic; antiinflammatory; antiarthritic; antiasthmatic;			
KW	antidiabetic; antiarteriosclerotic; vulnerary.			
XX				
OS	Homo sapiens.			
XX				
PN	WO200146392-A2.			
XX				
PD	28-JUN-2001.			
XX				
PF	21-DEC-2000; 2000WO-GB04963.			
XX				
PR	22-DEC-1999; 99GB-0030392.			
PR	07-APR-2000; 2000GB-0008713.			
XX				
PA	(OXFO-) OXFORD GLYCOSCIENCES UK LTD.			
XX				
PI	McKenzie EA, Stamps AC, Terrett JA, Tyson KL;			
XX				
XX	WPI; 2001-418056/44.			
DR	N-PSDB; AAH22673.			
DR				

XX	Novel homologs of heparanase, present in three splice variants, useful for identifying agents that modulate heparanase, useful in the treatment and/or prophylaxis of abnormal levels of heparanase -	
PT	Claim 1; Fig 3; 97pp; English.	
XX		
XX	The invention provides a homologue to heparanase which is present in three splice variants. The heparanase homologue polypeptide is useful in the treatment of a human or non-human animal or for use in diagnosis.	
CC	Vectors comprising the heparanase homologue polynucleotides are useful in the transformation or transfection of a prokaryotic or eukaryotic host.	
CC	The modulators of the polypeptide are useful in the manufacture of a medicament for the treatment and/or prophylaxis of a condition/disease associated with abnormal levels of the heparanase homologue, including cancer, central nervous system (CNS) and neurodegenerative diseases,	
CC	cardiovascular diseases such as restenosis following angioplasty and atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus, allografts, inflammatory diseases, arthritis, vascular restenosis,	
CC	tumour growth and progression, asthma, Alzheimer's disease, diabetic retinopathy, wound healing and inflammation. The polypeptide is also useful in diagnosis and research. The present sequence represents the	
CC	amino acid sequence of the smallest splice variant of the heparanase-like protein Hpa2 of the invention.	
XX		
SQ	Sequence 480 AA:	
Query Match 88.1%; Score 2270; DB 22; Length 480;		
Best Local Similarity 89.0%; Pred. No. 4.4e-221;		
Matches 439; Conservative 0; Mismatches 0; Indels 54; Gaps 1;		
Qy	1	GDRPLPVDRAAGLKEKTLILLDVSTKNPVTYNENFLSLQLDPSIIHDGWLDFLSKRL 60
Db	42	gdrplpvdraaglikektilldvstknprtvenflslqldpsiihdgwlflskrl 101
Qy	61	VTLAGLSPAFIRFGGKRTDFLQFNLRNPAKSRGGPGPDYILKNYEDDIVRSDVALDKQ 120
Db	102	vtlarglspafirfggkrtdfllqfnlrnpaksrpgpdyilknyed----- 149
Qy	121	KGCKIAQHPDVMLELQREKAAQMHVLVLLKEQFSNTYSNLILTEPNNTYRTHMGRVNGSQL 180
Db	150	-----epnnyrtmhgravnsgql 167
Qy	181	KDXYTQLKSLLOPIRIYSRASLYGNIGRPKNVIALLDGPMKVAGSTVDVATVWQHCVYID 240
Db	168	gkdyiqkllqpiriyraslygnigrprknvialldgfmkvagstvdatvqhcyid 227
Qy	241	GRVVKVMDFLKTRLLDLSDDQIRKIQKVVNTYTPGKKIWLGVVTTTSAGGTNNLSDSYAA 300
Db	228	grvkvmdflktrlldtsdqlrkikqvntytgpkkiwlegvvtttsaggtnnlsdsyaa 287
Qy	301	GFLWLTGLMLANQGDIVYIRHSFDDHGYNHLVDQNFNPLDPDYLLYLKRLIGPKVLAV 360
Db	288	gflwltglmlanqgdvvyirhsfddhgynhlvdgnfnpldpdyllslylkrllgpkvlav 347
Qy	361	HVAGLQRPGRVIRDKLRIYAHCTNHNHNHYVSGTTLFTINLHRSRKKIKLAGTLRD 420
Db	348	hvaglrkprgrvirdklriyahctnhhnhnyvrgsitlfiinlhrsrrkiklagtlrd 407
Qy	421	KLHVQYLLQPYGQGLKSKSVOLNGOPLVMVDDGTLPDLKPRPLRAGRTLVIPTVMTGFF 480
Db	408	klvhqyllqpyggeglksksvqlnggplvmvddgtlplkprplragrtlvipvmtgff 467
Qy	481	VKKNVNALACRYR 493
Db	468	vvkvnvalacryr 480
RESULT 11		
AAY97634		
ID	AAY97634 standard; Protein; 480 AA.	
XX		

AC AAY97634;
XX 20-APR-2001 (first entry)
DT Human heparanase, hnhp1 pn5 form, protein sequence.
XX
DE Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scarape;
KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
KW gene therapy; human.
XX
OS Homo sapiens.
XX
XX WO200100643-A2.
XX
XX 04-JAN-2001.
XX
XX 19-JUN-2000; 2000WO-IL00358.
XX
XX 25-JUN-1999; 99US-0140801.
XX
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
XX Pecker I, Michal I, Itzhaki H;
XX
XX WPI; 2001-137930/14.
DR N-PSDB; AAA91099.
XX
XX New polynucleotides and polypeptides that are distantly homologous to
PT heparanase, useful in wound healing, as well as in gene therapy
PT protocols for angiogenesis, restenosis, atherosclerosis, or
PT inflammation -
XX
XX Claim 10; Page 63; 67pp; English.
XX
XX This sequence represents a heparanase of the invention.
CC The heparanase DNA and protein sequences are useful in wound healing,
CC angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary
CC diseases, neurodegenerative diseases (such as Scarape, Alzheimer's
CC disease, and Creutzfeldt-Jakob disease) or viral infections. The
CC heparanase coding sequence is particularly useful in gene therapy.
XX
XX Sequence 480 AA;
SQ

Query Match 88.1%; Score 2270; DB 22; Length 480;
Best Local Similarity 89.0%; Pred. No. 4.4e-221;
Matches 439; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

QY 1 GDRPLPVDRAAGLKEKTLILLDVSTKNPVTNENFSLQLDPSIIHDGWLDFLSKRL 60
Db 42 gdrplpvdraaglkeltllldvstknprtvtnefnslqlpsiihdgwldfisskrl 101
QY 61 VTLARGSPALRGGRKRTDFLQONLRNPAKSRGGPGPDYLLKNYEDDIYRSDVALDKQ 120
Db 102 vtlargspalrfgrgrkrtdfllqonlrnpaksrpgpdpdyllknyed----- 149
QY 121 KGCKIAQHPDVMLELQREKAAQMLHLLVLLKEQFSNTYSNLIUTEPPNRYTMHGRAVNGSQL 180
Db 150 -----epnnrytmgravnsgql 167
QY 181 GKDYIQLKSLLOPRTIYRSASLYGNIGRPRKNVIALLDGPMKVGSTVDVATWQHGYID 240
Db 168 gkdyiqikslloprtiyrslaslygnigrprknvialldgfmkvagstvdvathqcyid 227
QY 241 GRVVKVMDFLKRLDLLDLSQIRKIQKVVNTYTPGKKIWLGVVTTGAGGTNNLSDSYAA 300
Db 228 grvvkvmdfllktrlldlsdqirkiqkvvntytpgkklwlgvvttsgaggtnnlsdsyaa 287
QY 301 GFLWNTLGLMLANOGIDWIRHSPFDHGYNHLVDQNFNPLPDYWLSLLYKRLIGPKVLAV 360
Db 288 gflwntlglmlanogidwlrhspfdhgynhlvdqnfnpdpdywlslllykrligpkvlav 347

QY 361 HVAGLQKRPGRVIRDKLRIYAHCTNHHNNHNYVRSITLFIINLHRSRKKIKIAGTLRD 420
Db 348 hvaglqrkprgrvirldkrlriyahctnhnhnyvrgsitlfiinlhrsrrkkikiagtlrd 407
QY 421 KLVHQYLLQPYGQSEGLKSKSVQNLNGOPLVMVYDDGTLPELKPRLRAGRTLVIPTVWGFF 480
Db 408 klvhyllqpygggglksksvqlngqplvmvddgtlpeikprprlragrtlviptvwmgff 467
QY 481 VVKNVNALACRYR 493
Db 468 vvkvnalacryr 480

RESULT 12
AAM99905
ID AAM99905 standard; Protein; 214 AA.
XX
AC AAM99905;
XX
DT 07-JAN-2002 (first entry)
XX
DE Human excretory related polypeptide SEQ ID NO 642.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antischizoid; antianemic; antithrombotic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antidiabetic; antidiabetic; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine;
KW excretory system.
XX
OS Homo sapiens.
XX
XX WO200155313-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01323.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.

PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.

PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 03-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465569/50.
XX N-PSDB; AA198878.
DR
DR Isolated nucleic acid molecule encoding excretory system antigen is
XX used in preventing, treating or ameliorating a medical condition -
XX
XX Claim 11; SEQ ID NO 642; 574pp + Sequence Listing; English.
XX
XX The invention relates to novel excretory system related human
XX polynucleotides (AA198567-AA199503) and the encoded proteins
XX (AA199594-AA199913) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy, especially
XX disorders related to the excretory system. The genes are isolated
XX from a range of human tissues disclosed in the specification. The
XX nucleic acids, proteins, antibodies and (ant)agonists are useful in
XX the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemia;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 214 AA;

Query Match 43.2%; Score 1112; DB 22; Length 214;
Best Local Similarity 100.0%; Pred. No. 3.5e-104;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 284 VTTSAGGTNNLSDSYAAGFLWTLGMLANQIGDIVRIHSFDDHGYNHLVDONFPLPDY 343
|||||

Db	5	vtsaggtgnlstdsaagflwnltgmlanggidvvirhsffdgynhlvdqgnfplpd	64
QY	344	WLSLYKRLIGPKVLAVHAGLQRPRGVRIRDKLRIVAHCTNHHNNHYRGSTITLFI	403
Dd	65	wsllykrligpkvlavhvagqlrkprgrvirdklriyahctnhnnhyrgstiltfi	124
QY	404	NLHSRKIKLAGTLRDKLVHQYLLOPYQGGLSKSVOLNGQPLVMVDDGTLP	463
Dd	125	nlhsrkkiklagtlrdklvhqyllpqyggelksksvglngqplvmvddgtlpelkprp	184
QY	464	Lragrtlvipptvmgffvvknvnalacryr	493
Dd	185	lrgrttvipptvmgffvvknvnalacr	214
RESULT	13		
AAM43704	ID	AAM43704 standard; Protein; 214 AA.	
XX	AC	AAM43704;	
XX	DT	24-OCT-2001 (first entry)	
XX	DE	Human bladder antigen, SEQ ID NO: 98.	
XX	KW	Human; bladder antigen; cytostatic; immunosuppressive; nootropic;	
KW	KW	neuroprotective; antiviral; antiallergic; hepatotrophic; antidiabetic;	
KW	KW	antiinflammatory; anticulcer; vulnerary; anticonvulsant; antibacterial;	
KW	KW	antifungal; antiparasitic; cardiant; gene therapy; cancer;	
KW	KW	immune disorder; cardiovascular disorder; wound healing; infection;	
XX	OS	neurological disease.	
XX	OS	Homo sapiens.	
PX	PN	WO200159064-A2.	
PD	PD	16-AUG-2001.	
XX	PF	17-JAN-2001; 2001WO-US01342.	
XX	PR	31-JAN-2000; 2000US-0179065.	
PR	PR	04-FEB-2000; 2000US-0180628.	
PR	PR	24-FEB-2000; 2000US-0184664.	
PR	PR	02-MAR-2000; 2000US-0186350.	
PR	PR	16-MAR-2000; 2000US-0189874.	
PR	PR	17-MAR-2000; 2000US-0190076.	
PR	PR	18-APR-2000; 2000US-0198123.	
PR	PR	19-MAY-2000; 2000US-0205515.	
PR	PR	07-JUN-2000; 2000US-0209467.	
PR	PR	28-JUN-2000; 2000US-0214886.	
PR	PR	30-JUN-2000; 2000US-0215135.	
PR	PR	07-JUL-2000; 2000US-0216647.	
PR	PR	07-JUL-2000; 2000US-0216880.	
PR	PR	11-JUL-2000; 2000US-0217487.	
PR	PR	11-JUL-2000; 2000US-0217496.	
PR	PR	14-JUL-2000; 2000US-0218290.	
PR	PR	26-JUL-2000; 2000US-0220963.	
PR	PR	26-JUL-2000; 2000US-0220964.	
PR	PR	14-AUG-2000; 2000US-0224518.	
PR	PR	14-AUG-2000; 2000US-0224519.	
PR	PR	14-AUG-2000; 2000US-0225213.	
PR	PR	14-AUG-2000; 2000US-0225214.	
PR	PR	14-AUG-2000; 2000US-0225266.	
PR	PR	14-AUG-2000; 2000US-0225267.	
PR	PR	14-AUG-2000; 2000US-0225268.	
PR	PR	14-AUG-2000; 2000US-0225270.	
PR	PR	14-AUG-2000; 2000US-0225447.	
PR	PR	14-AUG-2000; 2000US-0225757.	
PR	PR	14-AUG-2000; 2000US-0225758.	
PR	PR	14-AUG-2000; 2000US-0225759.	
PR	PR	18-AUG-2000; 2000US-0226279.	
PR	PR	22-AUG-2000; 2000US-0226681.	
PR	PR	22-AUG-2000; 2000US-0226686.	
PR	PR	17-NOV-2000; 2000US-0249208.	
PR	PR	17-NOV-2000; 2000US-0249207.	
PR	PR	08-NOV-2000; 2000US-0246613.	
PR	PR	08-NOV-2000; 2000US-0246611.	
PR	PR	08-NOV-2000; 2000US-0246610.	
PR	PR	08-NOV-2000; 2000US-0246609.	
PR	PR	08-NOV-2000; 2000US-0246532.	
PR	PR	08-NOV-2000; 2000US-0246524.	
PR	PR	08-NOV-2000; 2000US-0246525.	
PR	PR	08-NOV-2000; 2000US-0246526.	
PR	PR	08-NOV-2000; 2000US-0246527.	
PR	PR	08-NOV-2000; 2000US-0246528.	
PR	PR	08-NOV-2000; 2000US-0246478.	
PR	PR	08-NOV-2000; 2000US-0246477.	
PR	PR	08-NOV-2000; 2000US-0246475.	
PR	PR	08-NOV-2000; 2000US-0246476.	
PR	PR	08-NOV-2000; 2000US-0246477.	
PR	PR	08-NOV-2000; 2000US-0246478.	
PR	PR	08-NOV-2000; 2000US-0246479.	
PR	PR	08-NOV-2000; 2000US-0246480.	
PR	PR	08-NOV-2000; 2000US-0246481.	
PR	PR	08-NOV-2000; 2000US-0246482.	
PR	PR	08-NOV-2000; 2000US-0246483.	
PR	PR	08-NOV-2000; 2000US-0246484.	
PR	PR	08-NOV-2000; 2000US-0246485.	
PR	PR	08-NOV-2000; 2000US-0246486.	
PR	PR	08-NOV-2000; 2000US-0246487.	
PR	PR	08-NOV-2000; 2000US-0246488.	
PR	PR	08-NOV-2000; 2000US-0246489.	
PR	PR	08-NOV-2000; 2000US-0246490.	
PR	PR	08-NOV-2000; 2000US-0	

PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249285.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI N-PSDB; AAI64065.
DR WPI; 2001-514652/56.
DR N-PSDB; AAI64065.
XX
XX Forty five bladder related polynucleotides, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases -
XX
XX Claim 11: SEQ ID NO 98; 482pp + sequence listing; English.
PS
XX The invention relates to forty five novel bladder related
CC polynucleotides. The polynucleotides and the polypeptides that they
CC encode are useful in the diagnosis, treatment and prevention of:
CC cancer, particularly breast and ovarian cancer, and other cancers
CC of the adrenal gland, bone, bone marrow, breast, gastrointestinal
CC tract, liver, lung, or urogenital system; immune disorders such as
CC Addison's disease, allergies, autoimmune haemolytic anaemia,
CC autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple
CC sclerosis, rheumatoid arthritis and ulcerative colitis; cardiovascular
CC disorders such as myocardial ischaemias; wound healing; neurological
CC diseases such as cerebral anoxia and epilepsy; and infectious diseases
CC such as viral, bacterial, fungal and parasitic infections. Numerous
CC examples of each type of disorder are given in the specification.
CC The polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. The polynucleotides are
CC useful for chromosome identification. They are also useful as probes
CC for diagnosing or treating a disorder related to the female reproductive
CC system, particularly breast and/or ovary cancer. The present
CC sequence is a novel bladder antigen provided in the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIFO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 214 AA;

Query Match 43.2%; Score 1112; DB 22; Length 214;
Best Local Similarity 100.0%; Pred. No. 3.5e-104;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 VTTSAGTNNLSDSYAAGFLWNLTLGLMLANOGIDVIRHSFEDHGYNHLVDQNFNLPDY 343
Db 5 vttsagtnmlsdsyaagflwntlgmlanogldvirhsffdhgynhlvdqnfnpdy 64
QY 344 WLSLYKRLIGPKVLAVHVAGLQKRPGRVIRDKLRIYAHCTNNHHNNVVRGSLTFLII 403
Db 65 wlslykrligpkvlavhvaglqrkprgrvirdklriyahctnnhnhnyvrgsltlfll 124
QY 404 NLHRSRKKIKLAGTLDKLVHQLLQPYGQEGLSKSVQLNGOPLVMVDDGTLPKLKPRP 463
Db 125 nlhrrskkiklagtlrdklvghyllqpygqeglkksvqlngqplvmvddgtlpelkprp 184
QY 464 LRAGRTLVIPPTMGPFVVKVNNALACRYR 493
Db 185 lragrtlvipptcmgffvkvnnalacryr 214
RESULT 14
AAM24147
ID AAM24147 standard; Protein; 262 AA.
XX AC
XX AAM24147;
XX 12-OCT-2001 (first entry)
XX Human EST encoded protein SEQ ID NO: 1672.
XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition.
XX Homo sapiens.
XX OS
XX PN WO200154477-A2.
XX PD 02-AUG-2001.
XX 25-JAN-2001; 2001WO-US02687.
XX 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX WPI; 2001-476164/51.
DR N-PSDB; AAH98806.
XX Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX Claim 20; Page 1122-1123; 1275pp; English.
XX The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention.
XX
XX Sequence 262 AA;

Query Match 42.9%; Score 1106; DB 22; Length 262;
Best Local Similarity 98.2%; Pred. No. 1.9e-103;

Matches 216; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GDRPLPVDRAAGLKEKTLILLDSTKNPVTNENFLSLQDPSIIHDGWLDFLSSRL 60
Db 42 gdrplpvdraaglkektlilldvstknpvtvnenflslqldpsiihdgwlflsskrl 101

QY 61 VTLARGISPAFLRFGKRTDFLOFQNLNPNPAKSRGGPGPDYILKNYEDDIVRSDVALDKQ 120
Db 102 vtlargispaflrfggkrtdflofqnlnpnaksr99pgpdylknyeddivrsvaldkq 161

QY 121 KGCKIAQHPDVMLEQREKAAQMHLLVLLKEQFSNTYSNLILTEPNNYRTMHGRAVNGSQL 180
Db 162 kgckiaqhpdgmlprekaaqmhllvlkqfntsylntepnnyrtmhgravngsql 221

QY 181 GKDYIQLKSLLOPRTIYRSASLYGNIGRPRKNVIALLDG 220
Db 222 gkdyiqikslilqiriysraslygnivprknvialldg 261

RESULT 15

AAB08851
ID AAB08851 standard; Protein; 535 AA.

XX AC AAB08851;

XX DT 15-JAN-2001 (first entry)

XX XX A murine heparanase polypeptide.

Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
heparin-binding growth factor; cytokine; neurodegenerative plaque;
wound healing; infection; burn; angiogenesis; restenosis;
atherosclerosis; inflammation; neurodegenerative disease;
Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease.

XX OS Mus sp.

XX PN WO200052178-A1.

XX PD 08-SEP-2000.

XX PF 14-FEB-2000; 2000WO-US03542.

XX PR 01-MAR-1999; 99US-0258892.

XX PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX PA (FRIE/) FRIEDMAN M M.

XX PI Pecker I, Vlodavsky I, Feinstein E;

XX DR WPI; 2000-579289/54.

XX DR N-PSDB; AAAV5081.

XX XX New polynucleotides encoding a polypeptide having heparanase activity,
XX PT useful in wound healing and in gene therapy, particularly in treating
XX PT tumour, inflammation, autoimmunity, neurodegenerative diseases

XX PS Claim 22; Page 144-145; 152pp; English.

XX CC The present sequence represents murine protein with heparanase catalytic
XX CC activity. The heparanase (hpa) polynucleotide is useful in gene therapy,
XX CC particularly in treating tumour, inflammation or autoimmunity.
XX CC Particularly, the polynucleotide is useful in modulating the
XX CC bioavailability of heparin-binding growth factors, cellular responses
XX CC to heparin-binding growth factors (e.g. bFGF) and cytokines
XX CC (e.g. interleukin (IL)-8), cell interaction with plasma lipoproteins,
XX CC cellular susceptibility to certain viral and some bacterial and protozoa
XX CC infections, or disintegration of neurodegenerative plaques. The
XX CC polynucleotide is also useful in wound healing (e.g. thermal, chemical
XX CC or radiation burns), and in the treatment of angiogenesis, restenosis,
XX CC atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-
XX CC Straussler Syndrome or Creutzfeldt-Jakob disease), and some viral,

CC bacterial or protozoa infections.

XX Sequence 535 AA;

Query Match 36.7%; Score 944.5; DB 21; Length 535;
Best Local Similarity 40.0%; Pred. No. 1.5e-86;
Matches 211; Conservative 76; Mismatches 156; Indels 85; Gaps 9;

QY 22 LDVSTKNPVRTNENFLSLQDPSIIHD-GWLDFLSSKRLVTLARGLSPAFRLFGGKRTD 80

Db 33 lefytkrlrsvspflsitaslatdprflftlgsprlralarglspaylrfgtktdd 92

QY 81 FLQFQNLNPNPAKSRGGPGPDYILKNYEDDIVRSDVALDKQCKIAQHPDVMLEQREK 140

Db 93 flif----dpdkeptseersywkqvnhdicrsepv-----saavirklgvewp 137

QY 141 AQMHLLVLLKEQF-----SNTYS-----NLIL----- 161

Db 138 fq-eilllredyqkefkustysrsvdmlysfakcsqldlflgnallrtpdlrwnssna 196

QY 162 -----TEPNNYRTMHGRAVNGSQLGKDYIQLKSLLOPRTIYRSASL 202

Db 197 qlldycsskgyniswelgnepsfwkkaahlidqlgedfvelhklqr-safqnakl 255

QY 203 YGPNIGRPRKNVIALLDGFMKVAGSTVDATVWQHGYIDGRVVVKVMDFLKRLDLSDOI 262

Db 256 ygpdigprgktvkllrslfkaggevidstwhyylogriatkedflssdaldtflsv 315

QY 263 RKIQKVVNTYTPGKKIWLLEGVVTTTSAGGTNNLSDSYAAGFLWNLTLGMLANGIDVIRH 322

Db 316 qkilkvtkeitpgkkwligetssayggapllsntfaagfmldklgisagmgievvmrq 375

QY 323 SFFDHGYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHVAGLQKRPRGRVIRDKLRIY 382

Db 376 vffgagnyhlvdenfepldpywlsllfkklvgprvllsrvkpdp-----rsklrvy 426

QY 383 AHCTNHHNNHYVRGSIITFIINLHRSRKKIKLAGTLRDLKLVHQYLLQPYQGQGLSKSVQ 442

Db 427 lhctnvhypryqegdltyvlalnhtvkhklypppfrkpvdtlylkpsgpgdglsskvsq 486

QY 443 LNGQPLVMVDDGTLPELKPRLRAGRTLVPVPTMGFFVVKVNNALAC 490

Db 487 lngqilkmvdeqtlpalteklpagsalslpafsygfvirnakiac 534

Search completed: July 30, 2002, 08:14:02
Job time: 339 sec

us-09-836-461-2_copy_42_534.rag

Tue Jul 30 10:12:04 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:16:24 ; Search time 71.68 seconds
(without alignments)
660.882 Million cell updates/sec

Title: US-09-836-461-2_COPY_42_534
Perfect score: 2577
Sequence: 1 GDRRLPVDRAAGLKEKTLI.....PVTMGFFVKNVNALACRYR 493

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2270	88.1	480	2 JC7506	heparanase protein
2	302	11.7	521	2 T45608	hypothetical prote
3	155	6.0	190	2 T01953	hypothetical prote
4	111	4.3	1260	2 T14022	reverse transcript
5	109	4.2	986	2 F38229	sarcosine oxidase
6	109	4.2	986	2 AH3056	glutamate dehydrog
7	105	4.1	1020	2 T18342	protein-tyrosine k
8	101.5	3.9	629	2 C64180	cytochrome-c oxida
9	101.5	3.9	1125	1 JH0771	adenylate cyclase
10	101	3.9	535	2 S58740	hypothetical prote
11	101	3.9	2026	1 O1BY	probable glucosyl
12	100.5	3.9	582	2 A84206	hypothetical prote
13	100.5	3.9	630	2 C71374	probable membrane
14	99.5	3.9	455	2 S67627	hypothetical prote
15	99.5	3.9	1272	2 C50593	PRP8 protein homol
16	99	3.8	2403	2 T30875	tryptophan--trna l
17	98.5	3.8	335	2 E84992	endopeptidase Clp
18	97.5	3.8	845	1 G72079	ClpC proteinase [i
19	97.5	3.8	845	2 C86545	ATP-dependent Clp
20	97.5	3.8	845	2 B81590	DNA-directed RNA P
21	97.5	3.8	1517	2 BG1393	DNA polymerase III
22	96.5	3.7	969	2 D82895	probable helicase
23	96	3.7	763	2 B84432	SIP3 protein - yea
24	96	3.7	898	2 B4471	cytoplasmic aconit
25	96	3.7	1229	2 S42391	qsopa protein - Co
26	95.5	3.7	406	2 S68866	hypothetical prote
27	95.5	3.7	472	2 H90563	hypothetical prote
28	95.5	3.7	965	2 T22933	hypothetical prote
29	95	3.7	456	2 T24442	hypothetical prote

30	95	3.7	512	2 T11261	cytochrome-c oxida
31	95	3.7	557	2 D97210	extracellular neut
32	95	3.7	740	2 T51619	probable ethylene
33	94.5	3.7	460	2 G90554	p46-like (mycoplas
34	94.5	3.7	1280	2 B34087	hypothetical prote
35	94.5	3.7	1391	2 T20642	hypothetical prote
36	94.5	3.7	1397	2 E87998	protein F09C3.1 [i
37	94	3.6	625	2 G86855	glucose inhibited
38	94	3.6	902	2 D71079	hypothetical prote
39	94	3.6	3194	2 D71917	toxin-like outer m
40	93	3.6	320	2 D90478	conserved hypothet
41	93	3.6	484	2 C86264	protein kin-15 [im
42	93	3.6	488	2 I44330	protein-tyrosine k
43	93	3.6	849	2 S77217	phosphorylase (EC
44	93	3.6	894	2 T15769	hypothetical prote
45	92.5	3.6	328	2 AC1210	dTDP-D-glucose 4,6

ALIGNMENTS

RESULT 1
JC7506
heparanase protein 2a - human
C:Species: Homo sapiens (man)
C:Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 01-Dec-2000
C:Accession: JC7506
R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hircock, M.;
Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000
A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase fami
A:Reference number: JC7506
A:Accession: JC7506
A:Molecule type: mRNA
A:Residues: 1-480 <MCK>
A:Cross-references: GB:AF282885
C:Comment: This protein, a intracellular membrane-bound enzyme, has biological and th
therapies.
C:Genetics:
A:Gene: hpa2a
A:Map position: 10q23-10q24
C:KeyWords: heparin binding; membrane bound

Query Match	88.1%	Score	2270;	DB	2;	Length	480;
Best Local Similarity	89.0%	Pred. No.	7.5e-170;				
Matches	439;	Conservative	0;	Mismatches	0;	Indels	54;
Gaps	1;						
OY	1	GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHGDWDLFLSSKRL	60				
Db	42	GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHGDWDLFLSSKRL	101				
OY	61	VTLARGLSPAFIRFGGKRTDFLQFQNLNRPAPKSRGGPGPDYLLKNYEDDIVRSDVALDKQ	120				
Db	102	VTLARGLSPAFIRFGGKRTDFLQFQNLNRPAPKSRGGPGPDYLLKNYEDDIVRSDVALDKQ	149				
OY	121	KGCKIAQHPDVMLELQREKAAQMHVLLKKEQFSNTYSNLITPEPNNYRTHMGRAVNGSQL	180				
Db	150	-----EPNNYRTHMGRAVNGSQL	167				
OY	181	GKDYTLQKSLQPIRIYRSASLYGNIGRPNKVNIALDGFMKVAGSTVDVATWQHICYID	240				
Db	168	GKDYTLQKSLQPIRIYRSASLYGNIGRPNKVNIALDGFMKVAGSTVDVATWQHICYID	227				
OY	241	GRVVKVMDFLKTRLLDFTLSQIRKTKQVYNTYTPGKKIWLEGVVTTSAGTNNLSDSYAA	300				
Db	228	GRVVKVMDFLKTRLLDFTLSQIRKTKQVYNTYTPGKKIWLEGVVTTSAGTNNLSDSYAA	287				
OY	301	GFLWLTGLMLANQGDVIVIRHSFFDHGYNHLVDQNFNPLPDYWLSSLKYKRLIGPKVLAV	360				
Db	288	GFLWLTGLMLANQGDVIVIRHSFFDHGYNHLVDQNFNPLPDYWLSSLKYKRLIGPKVLAV	347				
OY	361	HVAGLQKRPGRGRVIRDKLURIYAHTCHTNHHNHYVRGSITFLFINLHRSRKKIKLAGTLRD	420				

Db 348 HVAGLQRPGRVIRDKLRIYAHCTNHHNNHNVYRGSITLFIINLHRSKKIKLAGTLRD 407

QY 421 KLHVOYLQPYQEGKSKSVOLNGOPLVMVDDGTLPELKPRLRAGRTLVIPTVTMGFF 480

Db 408 KLHVOYLQPYQEGKSKSVOLNGOPLVMVDDGTLPELKPRLRAGRTLVIPTVTMGFF 467

QY 481 VYKNNALACRYR 493

Db 468 VYKNNALACRYR 480

RESULT 2

T45608

hypothetical protein F13G24.30 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000

C:Accession: T45608

R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba

submitted to the Protein Sequence Database, December 1999

A:Reference number: 223009

A:Accession: T45608

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-521 <BEV>

A:Cross-references: EMBL:AL133421

A:Experimental source: cultivar Columbia; BAC clone F13G24

C:Genetics:

A:Map position: 5

A:Introns: 53/3; 66/1; 127/2; 177/1; 256/1; 319/2; 361/2; 394/3

A:Note: F13G24.30

Query Match 11.7%; Score 302; DB 2; Length 521;

Best Local Similarity 24.8%; Pred. No. 1.2e-15;

Matches 139; Conservative 73; Mismatches 183; Indels 166; Gaps 27;

QY 34 NENFLSLQD--PSIIFD-----GWLDFLS---SKRLVTLA-RGLSPAFLRFGGR 78

Db 22 DENFVCATLDWPP---HDKCNVDQCPWGYSSVINMDLTPRLTKAIFKPLRIRIGSSL 78

QY 79 TDFLQFQ--NLNRP-----KSRGGPGPDYLLKNEYDDIVRSOVALDKQKGIAPDVM 132

Db 79 QDOVIYDVGNLKTPCRPFQKMSG-----LFGFSKGLHMRWD-- 117

QY 133 LELOREKAAQMHV-----LLKEQFSNTYSNLILTEPNRYRTMHGRAVN---- 176

Db 118 -ELNSFLTATGAVVTGLNALGRHKLRKANGGAWDHINTODFLNTVTSKGVYDIDSWEF 176

QY 177 GSOL-----GKDYIQLKSLQPIRIYSRASLYGPNIGRPRKNVIALDGF- 221

Db 177 GNELSGSGVGASVSAELYGKDLVLKDVIN--KVYKNSWLHKPIVAP-----GGFY 226

QY 222 -----MKVAG-STVDVAVTQHCVIDGR-----VVKVMDPLKTRLLDLSQIRKIQ 266

Db 227 EQQWYTKLLEISGPSVVDVVT-HHIYNLGSGNDPALVKKIMD---PSYLSQVSKTFKDVN 282

QY 267 KVYNTVTPGKKIWLEGVWTSAG---GTMNLSDSYAAGFLWMLTGLMLANQIDVVIRH 322

Db 283 QTIQEHGP-----WASPWGESGAYNSGGRHSDTIDFSFWDQLGMSARHNTKYVCQ 338

QY 323 SFDFHGYNHLVDQNFPLPDYMLSLLYKRLIGPKVLAVHAGLQRPGRVIRDKRLTY 382

Db 339 TLVGGFYGLLEKGTFFVPNDYYSALLHRLMGKGVLAQTDG-----PPQLRYV 387

QY 383 AHCNTNHHNNHNVYRGSITLFIINLH-----RSRKKIKLAGTLRDKLV 423

Db 388 AHCCKG-----RAGVTLLILNLSQSDFTVSYSNGINVLNAESRKKSLDITLKRPF 441

QY 424 -----HOYLLQPYQEG-LKSKSVOLNGOPLVMVDDGTLPELKPRLRAGRT 469

Db 442 WIGSKASDGYLNREEHLTP--ENGVLRSKTYVLNGSKLKPATGDIPLSLEPILRSVNSP 499

QY 470 LVIPPTVTMGFFVYKNNALAC 490

Db 500 LNVPLSMFSFIVLPNFDASAC 520

RESULT 3

T01953

hypothetical protein T2L5.6 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 21-Jan-2000

C:Accession: T01953

R:Geisel, C.; Smith, A.; Le, T.

submitted to the EMBL Data Library, October 1998

A:Description: The sequence of A. thaliana T2L5.

A:Reference number: 214470

A:Accession: T01953

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-190 <GEI>

A:Cross-references: EMBL:AF096371; NID:g3695386; PID:g3695392

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 4

A:Introns: 36/2; 69/3

A:Note: T2L5.6

C:Superfamily: Arabidopsis thaliana hypothetical protein T2L5.6

Query Match 6.0%; Score 155; DB 2; Length 190;

Best Local Similarity 24.9%; Pred. No. 8.6e-05;

Matches 49; Conservative 32; Mismatches 70; Indels 46; Gaps 6;

QY 321 RHFSEFDHGVNHLVDQNFPLPDYMLSLLYKRLIGPKVLAVHAGLQRPGRVIRDKLR 380

Db 12 ROSLIGGNGYGLLNTNFTPNPDYISALYRQLMGKRALFTTFSGTK-----KIR 60

QY 381 IYAHCTNHHNNHNVYRGSITLFIINLHRSR---KKIKLAGTLRDKLVHVOYLLQPY----- 431

Db 61 SYTHCARQSK-----GITVLLMLNLDNTTVVAKVELNNSF--SLRHTKHKMSYKRASSQ 112

QY 432 ---GQEG-----LKSKSVOLNGOPLVMVDDGTLPELKPRLRAGRTLVIPT 473

Db 113 LFGGNGVITQREYHHTAKDGNLHSQTMLLNGNALQVNSMGDLPPIEPIHINSTEPIITA 172

QY 474 PYTMGFFVYKNNALAC 490

Db 173 PYSIVFVHNRNVVPAC 189

RESULT 4

T14022

reverse transcriptase homolog - slime mold (Dictyostellium discoideum) retrotransposab

C:Species: Dictyostellium discoideum

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000

C:Accession: T14022

R:Winckler, T.; Tschepke, C.; de Hostos, E.L.; Jendretzke, A.; Dingermann, T.

Mol. Gen. Genet. 257, 655-661, 1998

A:Title: Tdd-3, a transfer RNA gene-associated poly(A) retrotransposon from Dictyoste

A:Reference number: 217858; MUID:98265925

A:Accession: T14022

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1260 <WIN>

A:Cross-references: EMBL:AF002669; NID:g2558820; PID:g2558822; PIDN:AAC48324.1

Query Match 4.3%; Score 111; DB 2; Length 1260;

Best Local Similarity 19.7%; Pred. No. 4.1;

Matches 129; Conservative 89; Mismatches 204; Indels 232; Gaps 34;

QY 28 NPVRTVNEFNFLSLQIDPSIIHDGWL---DFL---SSKRLV-----TLARGLSP 69

Db 94 NGIGILNHNHNNKLSFIIEGRLLIISLIDKDTTRILAIYVAPQPKRKTLASTLNK 153

```
QY 70 AFLREGKRTDLOFQNL-RNPAKSRGGPGPDYLYKNYEDDIVRSDVALDKQKCKIAQH 128
Db 154 HENN-----QYHNTLSNPNKNDIIAGDFNCLDNHNTSND-----DQGNLTUOS 199
QY 129 PDVM---LELOEKAQAHLVLLKEO--FSNTYSNLILTEPNRYRTMHGRA-----VNGSQ 179
Db 200 PDEMATVIAIRISNNLMDQLNKRPTESRTHN---TNNLTRLRLRLRIYLNNSL 256
QY 180 LKQVYQLKSLQO----- 193
Db 257 INYSQLYLNLNLPKINDIPLSDHNFSLTFTLHNIOTNMRRWRLKSSILSMLKNID 316
QY 194 --IRYSR-----ASLY-----GNIGRPRKNVALLD 219
Db 317 FLNGYSRELNSHNSISFSQNLNLLNKIKOLYTEPQKQNDYNNKANI-----KNLSISLE 372
QY 220 -GFMKVGASTVDVATW-----QHC-----YIDGRVVK-VMDFLKTRLDLDTL 258
Db 373 TEFKDOAFATLSAINESKREBQKQELNNYCEETSLKYISARIKRNHDFINAVKDTQ 432
QY 259 SQOIKRIQVWYTYT-----PGKKIWLEGVVVTTSAGGTNNLSDSYAAAGFLW 304
Db 433 GRTINKQELIEEYKYVYVSNLYDYKEDDPSPHYEILENNVTTRDSTWDLNEFTSQEIL 492
QY 305 --LNTLGLMANOGIDVIRHSFFDH-----GYN-----HLVDQNF----- 337
Db 493 EVIKQLNPKHSPGPGIPNLFVITHKEKLAPILASAFNDTLRNPHLISKNYKEGLIITIP 552
QY 338 -----NLPDVLWSLLYK---RLIGPKVLAVHVGLOQRKPR---PGRVIRDKLRIY 382
Db 553 KKGDPKELNRRPITLANCIYKIHSLKLNRIIPILTKVINNOKGFVPGRFLHNIISI 612
QY 383 ACHTNHNNHNYVSGSTTLFIINLHRSRKKIKLAGTLRDKL--VHOYLLQPYGO-EGLKSK 439
Db 613 NELINYCNDRKINGIITLYI---SKLLTRSHTVQSQIITTHQHSNOYINLIMLLTK 667
QY 440 S---VQLNQCPLVMWDDGLP-ELKRPRLRAGRFLVIPVPTMGFFVKKVNAIA 489
Db 668 SEARIEINGRT-----TIPFEIK-RGVKQG-----DPLSPTLFVLV-IEALA 707

RESULT 5
F98229 sarcosine oxidase alpha chain PA5418 [imported] - Agrobacterium tumefaciens (strain C58,
C:Species: Agrobacterium tumefaciens
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: F98229
A:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorillo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A937359; PMID:11743194
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-986 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK89360.1; PID:g15159208; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L1574
A:Map position: linear chromosome
```

```
Query Match 4.2%; Score 109; DB 2; Length 986;
Best Local Similarity 18.2%; Pred. No. 4.1;
Matches 122; Conservative 69; Mismatches 211; Indels 268; Gaps 27;

QY 2 DRRLPVDRAAGLK-----EKTLLLDVSTKNPVRTVNEFLSLQLDPSIHDGWLDF 54
Db 362 DQRLDEARALGIEVLAGHSVNTAGRLVSSMTVGRGSGNKRKTAIDALVYVSGW--- 418
QY 55 LSKRLVTLRLGPSAFLRGKRTDFLOFQNLN-----LAFDANQRFLPDIIHVQNCVSGACNGTDLDLVAITAEAAAGG 96
Db 419 TFSVHLFSOSRGK---LAFDANQRFLPDIIHVQNCVSGACNGTDLDLVAITAEAAAGG 474
```

```
QY 97 PGPDYYLYKN-----YEDDIVRSDVALDKQKCKIAQH 130
Db 475 GSAFSGENARAWTGMIGAAEGAGGTGVKAFIDFQHDVCAKDIRLAVREG----- 526
QY 131 VMLELOREKAAQHLLVLLKEQFSNTYSNLILTEPNRYRTMHGRAVNGSOLGKDY---IQLK 188
Db 527 -----MHSV---EHKRETTNGMASDQGKMSNMHGLAIAASEALGRDLPKVGLT 571
QY 189 SILQO-----IRIYSRASLYGPNIGRPNVIALLDG-----FMKVAGS 227
Db 572 TFRQPTPTVFTGLINHSGALFDPTRKTPMHEE-ELAAGAVFEDYGNKRWKRAFFPRAGE 630
QY 228 TVDAVTHQHCYIDGRVVKVMDFLKTRLDLDTLSQIIRKIQKV-----NTYTPGKKIWLGE 282
Db 631 DMHEAINRECKTVRTSGVGFDASTLKGIEVVGPDAAKFLNLITYNAWDTLKPGRCRY--G 688
QY 283 VYVTSAGGTNNLSDSYAAAGFLWNTLGLMANOGIDVIRHSFFDHGYNHLVDONFPLPD 342
Db 689 IMTREDGFV-----YDDG-----VVGRLAEDRFHVTITTTGGAPRVLQHMDYLOTEFPD 737
QY 343 --YWLSLLYKR-----LIGPK---VLAVHVGLOQRKPR-----PGRV--- 374
Db 738 LNVWLTSATEQWAVIAVQGPKAREVIAPFVEGIDLSPFAFPHMAVAEGKFCGYPTRLFRV 797
QY 375 -----IRDKLRIYACHTNHNNHNYVSGSTTLFIINLH----- 406
Db 798 SFTGELGFEINVPADYGAAVWSAIRDTEAVGGC-----LYGTETMHLRAEKGYII 849
QY 407 -----RSRKKIKLAGTLRDKLHVQYLLQPYGOEGL 436
Db 850 VGQDGTGTTPDDAGLAWAVSKKTKDFVGIRGLKRLDLTRTKQLV-----GL 898
QY 437 KKSQVQLNGQPLVMWDDG---TLPELKPRLP----- 464
Db 899 KTK-----DRLTVPDEGGQIVTDPN-QPKPMTMLGHVTSAYSENGLHSIAFALVADGR 951
QY 465 -RAGRTLVIIP 473
Db 952 ARMGETLYIP 961

RESULT 6
AH3056 sarcosine oxidase alpha subunit [imported] - Agrobacterium tumefaciens (strain C58, D
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AH3056
A:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB25777; PMID:11743193
A:Accession: AH3056
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-986 <KUR>
A:Cross-references: GB:AE008689; PIDN:AAI44870.1; PID:g17742518; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: soxA
A:Map position: linear chromosome

Query Match 4.2%; Score 109; DB 2; Length 986;
Best Local Similarity 18.2%; Pred. No. 4.1;
Matches 122; Conservative 69; Mismatches 211; Indels 268; Gaps 27;

QY 2 DRRLPVDRAAGLK-----EKTLLLDVSTKNPVRTVNEFLSLQLDPSIHDGWLDF 54
Db 419 TFSVHLFSOSRGK---LAFDANQRFLPDIIHVQNCVSGACNGTDLDLVAITAEAAAGG 474
```

Db 362 DQRLDEARALGIEVLGHVSIVNTAGRLRVSSMTVGRNGSNKRKXIAIDALVWSAGW--- 418
QY 55 LSKRLVTIARGLSPAFLRFGKRTDFLQFQNLNR-----PAKSRGG 96
Db 419 TFSVHLFSOSRGK---LKFDAANQRLPDIHVQNCVSIAGACNGTDDLVIAIAEAAAGG 474
QY 97 PGDPYYLKN-----YEDDIVRSVALDKQKCKIAQHPD 130
Db 475 GSAFSGENARAWTGMGAEGAGEGTGVKAFIDFQHDVCAKDIRLAVREG----- 526
QY 131 VMLELOREKAAQMHVLLKQFSNTYSNLILTEPNRYTMHGRAVNSQLGKDY--IQLK 188
Db 527 -----MHSV---EHLKRTFTMGASDQCKMSNMHGLAIAEALGRDLPKVGLT 571
QY 189 SLLOP-----IRYSRASLYGPNIGRPRKNVIALLDG-----FMKVAGS 227
Db 572 TFPQPTVPTFTGLINHSRGALFDPTRKTPMHEE-ELAAAGAVFEDVGNKRAWFFPRAGE 630
QY 228 TVDAVTWQHCYIDGRVVKVMDFLKTRLLDLSQIRKIQKV-----NTYTGKKIWLGG 282
Db 631 DMHEAINRECKTVRTSGVDFDASTLGIKIEVVGPDAAKFLNLITNAWDTLKPGRCRY--G 688
QY 283 VVTTAGTNNLSDSYAAGFLMLNTGLMANQIDVIRHSFEDHGYNHLVDQNFNPLPD 342
Db 689 IMTREDGFV-----YDDG-----VVGRLAEDRFHTVTTGGAPRVLQHMEDYLQTEFPD 737
QY 343 --YWLSLLYKR-----LIGPK-----VLAVHVAGLQKRP-----PGRV--- 374
Db 738 LNVWLTSATQWAVIAVQGPKAREVIAPFVEGIDLSPEAFPHMAVAEGFCGVPTRLFRV 797
QY 375 -----IRDKLRIYAHCTNHHNHNHVGRSGITLFIINLH----- 406
Db 798 SFTGELGFEINVPADYGAAVSAIORTKTEAVGCC-----LYGTETWHILRAEKGYII 849
QY 407 -----RSRKKIKLAGTLRDLKLVHVOYLLQPYQOGL 436
Db 850 VGQDTCGTPTPDAGLAWAVSKKTDFTVGIRGLKRIDLTLTRGKQLV-----GL 898
QY 437 KRSVQLNGQLVMDVG-----TLPELKRPL----- 464
Db 899 KTK-----DRLTVPDGGQIVTDPN-OPKPMTMLGHVTSAYWSENGLHSIAFALVADGR 951
QY 465 -RAGRTLVIP 473
Db 952 ARWGETIYIP 961
RESULT 7
T18342
glutamate dehydrogenase (EC 1.4.1.2) precursor - Sauroleishmania tarentolae
C:Species: Sauroleishmania tarentolae
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-May-2000
C:Accession: T18342
R:Bringaard, F.; Striepecke, R.; Frech, G.C.; Freedland, S.; Turck, C.; Byrne, E.M.; Simps
Mol. Cell. Biol. 17, 3915-3923, 1997
A:Title: Mitochondrial glutamate dehydrogenase from Leishmania tarentolae is a guide RNA
A:Reference number: Z18878; MUID:97342629
A:Accession: T18342
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1020 <BRI>
A:Cross-references: EMBL:U31177; NID:g945092; PID:g945093; PIDN:AAB62735.1
C:Genetics:
A:Gene: GDH
C:Superfamily: glutamate dehydrogenase
C:Keywords: NAD; oxidoreductase
Query Match 4.1%; Score 105; DB 2; Length 1020;
Best Local Similarity 25.5%; Pred. No. 8.8;
Matches 42; Conservative 30; Mismatches 51; Indels 42; Gaps 7;
QY 73 RFGGKRTDFLOFQNLRNPAKSRGGP-----GPDYILKNYEDDIVRSDVA-----LDKQ 120

Db 656 KLGLKESEMRKFQT-----GGPDGDLGSNEVLRSKKKMGVMDISASLHDPNGIDRE 707
QY 121 KGCKIAQHPDVMLELOREKAAQMHVLLKQFSNTYSNLILTEPNRYTMHGRAV-NGSQ 179
Db 708 ELARLAHRLPLREFRSRK-----LSPEGF-----LVLTEDHNVKLPDGLTLVEDGSR 754
QY 180 LGKDYIOLKSLLOPIRTYRSRASLYGPNIGRPRKNVIALLDGFMKV 224
Db 755 LRNEFHFLK-----YSDADVFPVCGGRPRSVTLNVGREFLKV 791
RESULT 8
C64180
hypochemical protein H11056 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: C64180
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: C64180
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-829 <TIGR>
A:Cross-references: GB:U32786; GB:U42023; NID:g1574605; PIDN:AAC22721.1; PID:g1574609
Query Match 3.9%; Score 101.5; DB 2; Length 629;
Best Local Similarity 19.9%; Pred. No. 8;
Matches 54; Conservative 50; Mismatches 98; Indels 69; Gaps 9;
QY 33 VVENFLSLQDPSIIHDGWLDFLSKRLVTLARGLSPAFLRGGKRTDFLOFQNLRNPAK 92
Db 198 IEQAYLKILMDIDIFRDNFNLTVTVTKTIGGVSGSSE-----GSKLKDSTEINVFKNR 252
QY 93 SRGGPGPDY-----YLNKYED-----IVRSVDALDKQKCKIAQHPDVMLE 134
Db 253 ERLFLNPVYQKTEVEFEIKNYEDSGSKWKYTOVLIDLGEKILLEKDGFKYVHYPN--- 308
QY 135 LOREKAAQMHVLLKQFSNTYSNLILTEPNN--YTMHGRAVNSQLKDKYIQLKSLQ 192
Db 309 -----AQMTSIVKFSQDONLSKEIITYEYSHKVYRTTNAQSSIRSKIEDLYSIKNGIV 362
QY 193 PRIYSRASLYGPNIGRPRKNVIALL-----DGMKVAGSTV-----D 230
Db 363 STB-----YIPQKGNAGNLIEFYNASNKDMFMSDLMLIKEKNKYFYLOKVTMLWD 415
QY 231 AVTWOHC-----YIDGRVVKVMDFLKTRLLD 256
Db 416 DIOYNLNLKEGGYIDFKNGKKEALLRIID 446
RESULT 9
JH0771
protein-tyrosine kinase (EC 2.7.1.112), receptor type hyk precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Jun-1993 #sequence_revision 10-Jun-1993 #text_change 21-Jul-2000
C:Accession: JH0771
R:Horita, K.; Yagi, T.; Kohmura, N.; Tomooka, Y.; Ikawa, Y.; Alzawa, S.
Biochem. Biophys. Res. Commun. 189, 1747-1753, 1992
A:Title: A novel tyrosine kinase, hyk, expressed in murine embryonic stem cells.
A:Reference number: JH0771; MUID:93129253
A:Accession: JH0771
A:Molecule type: mRNA
A:Residues: 1-1125 <HOR>
A:Cross-references: GB:D13738; NID:g220439; PIDN:BAA02883.1; PID:g220440
C:Genetics:
A:Gene: hyk

C:Superfamily: protein-tyrosine kinase, receptor type tie; EGF homology; fibronectin type C; Keywords: Anp; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho F;1-22/Domain: signal sequence #status predicted <Sig>
F;23-1125/Product: protein-tyrosine kinase, receptor type hyk #status predicted <MAT>
F;135-137/Region: immunoglobulin homology <IM1>
F;212-252/Domain: cell attachment (R-G-D) motif
F;256-299/Domain: EGF homology <EG1>
F;303-341/Domain: EGF homology <EG2>
F;303-341/Domain: EGF homology <EG3>
F;365-427/Domain: immunoglobulin homology <IM2>
F;448-526/Domain: fibronectin type III repeat homology <FN3A>
F;541-625/Domain: fibronectin type III repeat homology <FN3B>
F;638-720/Domain: fibronectin type III repeat homology <FN3C>
F;752-773/Domain: transmembrane #status predicted <TM>
F;823-1100/Domain: protein kinase homology <KIN>
F;831-839/Region: protein kinase ATP-binding motif
F;140,158,400,439,465,559,649,691/Binding site: carbohydrate (Asn) (covalent) #status F;856,873,965/Active site: Lys, Glu, Asp #status predicted

Query Match 3.9%; Score 101.5; DB 1; Length 1125;
Best Local Similarity 21.3%; Pred. No. 19;
Matches 80; Conservative 43; Mismatches 118; Indels 135; Gaps 19;
Qy 13 GLKEKTLILLDVSKNPVRTVNEF-----LSLQIDPSIIHD--GWLDFL 55
Db 701 GLEPETTHVDIFAEENIGSSNPAPSFHELRLPHSPASADLGKGDATHSPHWGWNDA 760
Qy 56 SSKRLVTLARGLSPAFL-----RFGKRTDFLOFNLR-NPA-----KSRGG 96
Db 761 SPCLL-----AFLIMLQLKRANVORRMAQAFQNVREPAFQNSGTLALNRKANN 811
Qy 97 PGPDY-----LK-NYEDDIVRSDVALDKQG----- 122
Db 812 PDPTIYPVLDMNDIKFDQVIGEGNFQGVKARIKKDGLRMDAAIKRMKEYASKDDHRDFA 871
Qy 123 -----CKIAQHPDVMLELQR-ERAAOMHLVLLKEQFSNTY-----SNLIITEPN-NYRT 169
Db 872 GELEVLCKLGHHPNIINLLGACEHGYLYLAIEYAPHGNLLDFLGKSRVLETDPAFCHQ 931
Qy 170 MHGRAVNSQL-----GKDYIQLKSLLOPIRIYSRASLYGPNIGRPKNVIALDGL 220
Db 932 QYSTSLSSQQLHFEADVARGMDVLSQKFTHR-DLAARNILVG-----ENYIAKID 983
Qy 221 FMKVAGSTVDVTHQCHYIDGRVVKVDFLKTR--LDLTSLDQIRKKIKVNTVTPGKKI 278
Db 984 FGLSRGQEV-----YVKRTMGRPLVRWMAIESLN-----YSVYTTNSDV 1022
Qy 279 WLEGVW---TTSAGGT 291
Db 1023 WSYGVLLWEIVSLGGT 1038

RESULT 10
S58740
Cytochrome-c oxidase (EC 1.9.3.1) chain I - yeast (Hansenula wingei) mitochondrion
C:Species: mitochondrion Hansenula wingei
C:Date: 19-Mar-1997 #sequence_revision 09-May-1997 #text_change 07-Dec-1999
C:Accession: S58740
R:Seikito, T.; Okamoto, K.; Kitano, H.; Yoshida, K.
Curr. Genet. 28, 39-53, 1995
A:Reference number: S58740; MUID:96022424
A:Accession: S58740
A:Molecule type: DNA
A:Residues: 1-535 <SEK>
A:Cross-references: EMBL:D31785
A:Note: the authors translated the codon CAA for residue 54 as Gly
C:Genetics:
A:Gene: cox1
A:Genome: mitochondrion
A:Genetic code: SGC3
A:Introns: 241/3

C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
C:Keywords: chromoprotein; copper; electron transfer; heme; iron; lipoprotein; magnes
ductase; respiratory chain; transmembrane protein
F;11-458/Domain: cytochrome-c oxidase chain I homology <CO1>
F;63,379/Binding site: heme a iron (His) (axial ligands) #status predicted
F;242,291,292/Binding site: copper (His) #status predicted
F;242-246/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
F;246/Binding site: oxygen (Tyr) #status predicted
F;320/Binding site: myristate (Lys) (covalent) #status predicted
F;369/Binding site: magnesium (His) (shared with chain II) #status predicted
F;377/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 3.9%; Score 101; DB 2; Length 535;
Best Local Similarity 22.2%; Pred. No. 6.9;
Matches 68; Conservative 46; Mismatches 102; Indels 90; Gaps 19;
Qy 194 IRIYSRASLYGPNIGRPKNVIALDGLGFMKVGASTVDVTHQCHYIDGRVVKVMDFLKT 252
Db 260 VSTSKRPVGE-----ISWYAMASIA--FLGLVWSHHMYIVG-----LD-ADT 302
Qy 253 RLDTSLDQIRKKIKVNTVTPGKKIWLGVVTTTSAGTNLSDS--YAAGFLWLNTLG- 309
Db 303 RAYFTSSTMVIAVPTGIKFS-----WL-----ATLYGGSIRLAVPMLYAI AFLFLTIGG 353
Qy 310 -----MLANQGDIVIRHSFFDPHGYNHLV---DQNFNPDPYV-----LSLLYKR----- 351
Db 354 LTGVALANASLDVAFHDFTYVYVGHFYVLSMGAIFSLFAGYYWSPQILGLYFNERLAI 413
Qy 352 -----LIGPKV--LAVHVAGLQKRPGRVIRDKLRIYAHCTNHHNNHVRG----- 396
Db 414 QFWLIFGAVNIFPMPHFLGLQGHPR-----RIDPDYDAYAGWNYVSSIGSVIAII 464
Qy 397 SITLFIINLHRSRKKIKLAGTLRDKLVHQ---YLLQP-----YGOEGLKSKSVQ--L 443
Db 465 SLALFYIYID-----QLINGLTNNKIDNVVYSKAPDVFESNTIFANNKSASIEFL 519
Qy 444 NGQPLV 449
Db 520 NSPPAI 525

RESULT 11
OYBY
adenylate cyclase (EC 4.6.1.1) - yeast (Saccharomyces cerevisiae)
N:Alternate names: ATP pyrophosphate-lyase; protein J1401; protein YJL005W
C:Species: Saccharomyces cerevisiae
C:Date: 28-Dec-1987 #sequence_revision 08-Sep-1995 #text_change 21-Jan-2000
C:Accession: S56776; S56775; A24776; S05828; S55183
R:To Van, D.; Perea, J.; Jacq, C.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S56776
A:Accession: S56776
A:Molecule type: DNA
A:Residues: 1-1823 <DEH>
A:Cross-references: EMBL:Z49280; GSPDB:GN00010; MIPS:YJL005W
R:de Haan, M.; Grivell, L.A.; Smits, P.H.M.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S56771
A:Accession: S56775
A:Molecule type: DNA
A:Residues: 673-2026 <ZAG>
A:Cross-references: EMBL:Z49280; GSPDB:GN00010; MIPS:YJL005W
R:Kataoka, T.; Broek, D.; Wigler, M.
Cell 43, 493-505, 1985
A:Title: DNA sequence and characterization of the S. cerevisiae gene encoding adenyla
A:Reference number: A24776; MUID:86079531
A:Accession: A24776
A:Molecule type: DNA
A:Residues: 1-261, 'L', 263-547, 'L', 549-591, 'H', 593-708, 'I', 710-961, 'P', 963-1387, 'S', 13
A:Cross-references: EMBL:M12057; NID:g171359; PIDN:AAA34549.1; PID:g171360
A:Note: the authors translated the codon TTA for residue 262 as Ser. ACG for residue
2 as Leu, TCA for residue 1566 as Ala, GAG for residue 1659 as Ala, GGT for residue 1

R;Masson, P.; Lenzen, G.; Jacquemin, J.M.; Danchin, A.
Curr. Genet. 10, 343-352, 1986
A:Title: Yeast adenylate cyclase catalytic domain is carboxy terminal.
A:Reference number: S05828; MUID:88165073
A:Accession: S05828
A:Molecule type: DNA
A:Residues: 1042-1426, 'D', 1428-1460, 'T', 1462-1955, 'V', 1957-2008, 'M', 2014-2015, 'T', 'T', 'N', 'F', 'Y', 'K', 'W', 'L', 'R', 'T'
A:Cross-references: EMBL:X03449; NID:93487; PIDN:CAA27175.1; PID:93488
R;de Haan, M.; Smits, P.H.M.; Grivell, L.A.
submitted to the EMBL Data Library, May 1995
A:Reference number: S55183
A:Accession: S55183
A:Molecule type: DNA
A:Residues: 673-2026 <DEW>
A:Cross-references: EMBL:X87611; NID:9854567; PIDN:CAA60917.1; PID:9854568
C:Genetics:
A:Gene: SGD:CYR1; MIPS:YJL005W
A:Cross-references: SGD:S0003542; MIPS:YJL005W
A:Map position: 10L
C:Superfamily: yeast adenylate cyclase; leucine-rich alpha-2-glycoprotein repeat homolog
C:Keywords: CAMP biosynthesis; duplication; phosphorus-oxygen lyase; tandem repeat
F:669-1343/Region: leucine-rich 23-residue repeats
F:1065-1087/Domain: leucine-rich alpha-2-glycoprotein repeat homolog <LRR>
F:1610-2026/Domain: catalytic status predicted <CAT>
F:1664-1749/Domain: yeast adenylate cyclase catalytic domain homolog <YACC>

Query Match 3.9%; Score 101; DB 1; Length 2026;
Best Local Similarity 20.1%; Pred. No. 50;
Matches 120; Conservative 80; Mismatches 201; Indels 196; Gaps 30;

QY 12 AGLKE-KTLLLDVSTKPV---RTVNEFLSLQDPSI-----I 47
DB 903 AGFVELKNQLDLSNFMHYPVINYCTNLQIDLSNKIQSLPQSTKYLVKLAKMNL 962

QY 48 HDGWLDFSSKRLVTLARGLSPAFRFGKRTDFQFQNL----- 87
DB 963 SHKNLNFIDGSEMDLRTFLNLRYNRISSIKTNASLQNLFLTDNRISNEDTLPKLRAL 1022

QY 88 ---RNPASRGSGPPDYLYKNYEDDIDVSDVADKQKCKTAQHPDVMLE---LQREKA 140
DB 1023 ETQENPITSIS--FKDFYKPN-----MTSLTLNK---AQLSSIPGELLTKLSFLEKLEL 1071

QY 141 AQHVLVLLKEQFSNTYSNLILTEPNN-----YRT--MHGRAVNGSOLGKDY 184
DB 1072 NONNLRFPQESKLTLYLVLSVARNKLEYIPPELSQLKSLRTLDLHNSNIRDFVDGMEN 1131

QY 185 IQLKSLQPIRIYSRASL-----YGNIGRPRKNVIALDGFPMKVAGSTVDVATWQ 235
DB 1132 LEITSLSNLSNAPGSSLENSFVHMSYGSKLSK-----SLM--FFIAADNQFDAMWP 1183

QY 236 ---HCYIDGRVVKVMDPLKTRLLDLSQIRKIQKVNTVTPGKKIWLGVVTTTSAGTNN 293
DB 1184 LFNCFYN---LKVNLISYNNFSDV---SHMKLESITELYSGNKL-----TTLSGDT-- 1229

QY 294 LSDSYAAGFLW--LNTLGLMANGIDV-----VIRHSEDFHGYNHLVDQNFNLPDY-W 344
DB 1230 -----VLKWSLSKTLMLNSQMLSLPAELSNLSQSVDFVGANQKYNISNYHVDNNW 1282

QY 345 LS---LLYKRLIGP-----KVLAVHVGLOKRPGRVI 375
DB 1283 RNNKELKYLNFSGNRRRFEIKSFISHDIDADLSLTLPQLKVLGLMDVLTNTTKVPDENV 1342

QY 376 RDLIRYAHCTNNH-----NVV-----RGSITLFTINLHRSRKK----- 411
DB 1343 NFRLTASTINGMRGVADTLGQRIYVSSROVTFEFGNDECELLCHDSKNQADYG 1402

QY 412 ---IKLAGTLRDK-LVHQYLLQPYGQE-----GLKSKSVQLNGO---PLWVDDG 454
DB 1403 HNISRIVRDIYDKLIRQ--LERYGDETDDNKTALRFSFLQNLKNGMLNSVDNG 1457

RESULT 12

A84206
hypothetical protein porA [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
A:Accession: A84206
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: A84206
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-582 <STO>
A:Cross-references: GB:AE004437; NID:g10580081; PIDN:AAG19013.1; GSPDB:GNO0138
C:Genetics:
A:Gene: porA
C:Superfamily: Halobacterium halobium 2-oxoacid ferredoxin oxidoreductase; 2-oxoacid

Query Match 3.9%; Score 100.5; DB 2; Length 582;
Best Local Similarity 21.2%; Pred. No. 8.5;
Matches 69; Conservative 49; Mismatches 110; Indels 97; Gaps 14;

QY 8 VDRAAGLKEKTLILDVSTKNPVRVTNENFLSLQDPSIHDGWLDFLSSKRLVTLARGL 67
DB 58 VDQVQSVDRDLILIALTE---RTVDENLDELHADSIIIVD----- 95

QY 68 SPAFLRFGKRTDFLQFQNLNPAKSGGPGDYLYKNYEDD---IVRSDVALDKQKC 123
DB 96 -----GDRTEPADFE---SPAEV---TGLDIPLKDLAEDAGGATMRNIVALGAV--C 139

QY 124 KIAQHPDVMLEQREKAAQMHVLLKKEQFSNTYSNLILTEPNNRYTMHGRAVNGSOLGKD 183
DB 140 AVADFPFIEENDESLEK-----RFGSGEQI---ENNKQ-----AARLGAE 177

QY 184 YI--QLKSLQPIRIYSRASLYGNIGRPRKNVIALDGFPMKVAGSTVDVATWQHCYIDG 241
DB 178 YVAEEFEDVTLPELETDDYVLLNGDEAIGMAIAAGCRFYAGYPTPAT----- 229

QY 242 RVKVMDFLTKRL-----LDTLSDQIRKIQKVNTVTPGKKIWLGVVTTTSAGTNNLS 295
DB 230 ---DVMEYLTGRIFQFGHVVQAEDELAALNLAARAGAR----SWTATSGPGIDLMS 282

QY 296 DSYAAGFLWLTNGLMANGIDVVI 320
DB 283 E-----TFGLVATSETPLVI 297

RESULT 13
C71374
probable glucose inhibited division protein A (gida) - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 26-Aug-1999
A:Accession: C71374
R;Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770
A:Accession: C71374
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-630 <COL>
A:Cross-references: GB:AE001189; GB:AE000520; NID:g3322293; PIDN:AAG65038.1; PID:g332
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0044
C:Superfamily: gida protein

Query Match 3.9%; Score 100.5; DB 2; Length 630;
Best Local Similarity 21.3%; Pred. No. 9.6;
Matches 77; Conservative 52; Mismatches 129; Indels 103; Gaps 20;

QY 177 GSOLGK--DYTLQKSLLOPIRIYSRASLYGNIGPRKNVIALLDGFMKVGASTVDATW 234
DB 68 GEMGKFAACMIQ-----YRLNKR--GPAVQAPRQADKFL--YAKVKYITLECTQH 118
QY 235 QRCYIDGRVWVMDFLKTRLDLTSDQIRKIQKVNTVTPGKKIWLEGVWITSAGTNNL 294
DB 119 LHLQD-----FVVDVCSNTTDAGVYAGAAHVVT--ARGRRISARAVLTG----- 166
QY 295 SDSYAAGFLWL-----NTLGLMANQDIDVIRHSFFDHG-----YNHLVDQNF 337
DB 167 --TFMEGRVYIGEYAEPRGEGHAAEGALRRKKGFQMGRLKTGTTPARVLRSVD--- 221
QY 338 NPLPDYWLSLVKR----LIGPKVLAVHAGLQRPGRVIRDKLRIYAHCTNHHNHY 393
DB 222 -----LSVMEKQADAIMRPFSPA-HV-----EINRHADCYINTNERTHOL 263
QY 394 VRGSITFLIINLHRS---RKKIKLAGT-----LRDKLVHQLLOPYGOEGLKS 438
DB 264 IRE-----NFRSPFSGRIKAVGTRYCPSIEDKVRKFPDRIRHOLYIEP---EGLDT 313
QY 439 KSVQLNGOPLVWDD-----GTLPELK-----PRPLRAGRTLVPVPTMGF-FVYKNVNA 487
DB 314 BELXINGLSCLPEDIQDEMIRTIPGMEAVITRPAYAVDVAVLPVQLGIDLQTKRVSG 373
QY 488 L 488
DB 374 L 374

RESULT 14
S67627
Probable membrane protein YDL091c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein D2408
C:Species: Saccharomyces cerevisiae
C>Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 21-Nov-1997
C:Accession: S67627; S67633; S67426
R:Wambutt, R.; Wedler, H.; Wedler, E.; Scharfe, M.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S67608
A:Accession: S67627
A:Molecule type: DNA
A:Residues: 1-455 <WAM>
A:Cross-references: EMBL:Z74139; NID:gl4311121; PID:e253025; PID:gl4311122; MIPS:YDL091c
A:Experimental source: strain S288C
R:Ballesta, J.P.G.; Remacha, M.; Soler-Mira, A.; Jimenez, A.; Garcia-Cantalejo, J.M.; Bo
submitted to the Protein Sequence Database, July 1996
A:Reference number: S67629
A:Accession: S67633
A:Molecule type: DNA
A:Residues: 1-455 <BAL>
A:Cross-references: EMBL:Z74139; NID:gl4311121; PID:e253025; PID:gl4311122; MIPS:YDL091c
A:Experimental source: strain S288C
R:Boskovic, J.; Saliz, J.E.; Soler-Mira, A.; Garcia-Cantalejo, J.; Revuelta, J.L.; Jimine
submitted to the EMBL Data Library, February 1996
A:Reference number: S67406
A:Accession: S67426
A:Molecule type: DNA
A:Residues: 1-455 <BOS>
A:Cross-references: EMBL:X95644; NID:gl199535; PID:e223196; PID:gl199556
C:Genetics:
A:Map position: 4L
A:Note: YDL091c
C:Keywords: transmembrane protein
F:59-75/Domain: transmembrane #status predicted <TM>

Query Match 3.9%; Score 99.5; DB 2; Length 455;
Best Local Similarity 21.3%; Pred. No. 7.1;

Matches 95; Conservative 59; Mismatches 145; Indels 147; Gaps 22;

QY 69 PAFLRFGKRTDFLOFNLRNPAKSRGGPGPD-----YILKNYEDDVRSDV 115
DB 27 PADL--NGRTED--QNSNTNEPTQSRDGLKSLILHFLQAPLIVLYLLNF---IVRSR 79
QY 116 ALD-----KQCGCKIAQHPDVMLELQREKAAQMHVLLKQFSNTYSNLITPEPN 166
DB 80 LLKPLLLRHGFQYRKHNRLLDH-----SSQLHRL--ENLENAQAQVTCSEGN- 125
QY 167 YRTMGRVAVGSOLGKDYIQLKSLLOPIRIYSRASLYGNIGPRKNVIALLDGFMKVG 226
DB 126 -----GNDDGNTSTSTNNSSGVQ-----PSFGSLYINPENGTFKSII--MONSTELLD 174
QY 227 STDVAVTWHQCHYIDGRVVKVMDFLKTRLDLTSDQIRKI---QKVNTVTPGKKIWLEGV 283
DB 175 ACSEQVKEGVIY-----LHDPDLLNMDVYVVKILCSEAFVNMIRKQVQLLWYGD 223
QY 284 VTTSGAGTNNLSDSYAAGFLWNTLGLM---ANQGDIVVIR-----HSFFDH 327
DB 224 VTTSEG----LQVSNALKIROYPPLGLIISLKAERKIELIARVEGSGISNYKAODLEAFSK 279
QY 328 GYNHLVDQNFNPLPDYWLSSLVKRLIGPKVLAVHAGLQRPGRVIRDKLRIYAHCTN 387
DB 280 NYSRLIQ-----LROORONIEMQRLI-----RQOODSR-YODSLR----- 313
QY 388 HHNNHVVRGSITFLIINLHRSRKKIKLAGTLRD-----KLHVHOLL-----QPYG 432
DB 314 -----RDOQRESERLEOTQREOMERHOREHORIENQWLLWRKSQLEPSS 356
QY 433 QEGLKSKSVQL-NGQPLVWMDGTLP 457
DB 357 DKDASKVAIRLENGQRLVRKFDASLP 382

RESULT 15
C90593
Hypothetical protein MYPUL_6510 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: C90593
R:Chambaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer,
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma p
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: C90593
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1272 <KUR>
A:Cross-references: GB:AL445566; PID:gl4090066; PIDN:CAC13824.1; GSPDB:GN00153
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYPUL_6510
A:Genetic code: SGC3

Query Match 3.9%; Score 99.5; DB 2; Length 1272;
Best Local Similarity 18.6%; Pred. No. 33;
Matches 85; Conservative 69; Mismatches 129; Indels 173; Gaps 23;

QY 15 KEKTLIL--LDVSTK-----NPVRTVNENFLSLQL-----DPSI 46
DB 781 NKSITTPKDIQVKYKIASNPQYNLKQNSINVEFVEKNGHSQKVTYRLNGFKKFDAS- 839
QY 47 IHGWLDFLSSKRLV-----TLARGLSPAFLRFGGKRTDFLOQNLRNPAKSRG 95
DB 840 ----WFDVYAKGNILKONPRDIKQDKTHFKFIANTF---GTSKSFYFKFNQTSKE--- 888
QY 96 GPGPDYILKNYE---DDIVRSVDALDKQCKIAQHPDVMLELQREKAAQMHVLLKQF 152
DB 889 -----LLKHYNVSDNIEYKNVAGE-----THFDVKFTRKSSSSETNIV-IRKVF 933
QY 153 SNTYSNLILTEPNY-----RTMHGRAVNGSOLGKDYIQ-----LKSL 190

Tue Jul 30 10:12:06 2002

```

| :||| : | | : | : || :||
Db 934 SGFKDLIFTDEQYDDEDTREHRSIAITTKMPRNQFIQKIVEARQNNDLDALVDYFTNL 993
QY 191 LQPIRIYSRA---SLYGNIGRPRKNVIA-----LL 218
Db 994 IAPFRISNETLSFSIRTQNGSPRKNKGAHREDIADQTAEOFTILHRLSNLVTNKEKNIR 1053
QY 219 DGFMKVAGSTVDVWQHCHYIDGRVVKVMDFLKTRLL--DTLSQIRKIQKVVVWYTPGK 276
Db 1054 DHIREWLNETVFAIRTSH-----LKTSEEFKNEFLAKTSSERIELIKKYDYIP-- 1104
QY 277 KIWLEGVVTTSAGTNNLS-----DSYAAGF--LWLNLTGLMLANQIGIDVWVRH 322
Db 1105 ----EG-YTLNINYTHNITSKYKYDNOVERKENKLSDFEILYIITKGDKSSQ-----RK 1153
QY 323 SPFDHG---YNHLVDQNFNPL-----PDYW 344
Db 1154 LILNYGQRIYFNHPKANQLNPLFIKRDFLKYKDEYW 1189
```

Search completed: July 30, 2002, 08:16:27
Job time: 344 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:31:27 ; Search time 36.14 Seconds
(without alignments)
528.189 Million cell updates/sec

Title: US-09-836-461-2_COPY_42_534
Perfect score: 2577
Sequence: 1 GDRRLPVDRAAGLKEKTLI.....PVTMGFFVKNVNALACRYR 493

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101.5	3.9	629	1 T3MH_HAEIN	P71366 haemophilus
2	101	3.9	2026	1 CYAA_YEAST	P08678 saccharomyc
3	100.5	3.9	630	1 GIDA_TREPA	O83084 treponema p
4	98.5	3.8	335	1 SYW_BUCAI	P57602 buchnera ap
5	97.5	3.8	845	1 CLPC_CHLPI	Q928a6 chlamydia p
6	97.5	3.8	1517	1 RPOC_CAMJE	O9p130 campylobact
7	96.5	3.7	969	1 DP3A_UREPA	Q9p974 ureaplasma
8	96	3.7	1229	1 SIP3_YEAST	P38717 saccharomyc
9	94.5	3.7	682	1 SNK_RAT	Q9r012 rattus norv
10	94.5	3.7	685	1 SNK_HUMAN	Q9nyy3 homo sapien
11	94	3.6	625	1 GIDA_IACLA	Q9cej4 lactococcus
12	93.5	3.6	947	1 LKTA_PASSP	P55123 pasteurella
13	93	3.6	488	1 K115_CAEEL	P34891 caenorhabdi
14	93	3.6	849	1 PHSQ_SYNT3	P73511 synechocyst
15	92.5	3.6	682	1 SNK_MOUSE	P53351 mus musculus
16	92.5	3.6	690	1 HELS_METTH	O26901 methanobact
17	92.5	3.6	1616	1 VITI_CAEEL	P55155 caenorhabdi
18	92.5	3.6	2241	1 TEGU_HCMVA	P16785 human cytom
19	92	3.6	461	1 GPI3_YEAST	P32363 saccharomyc
20	92	3.6	534	1 COX1_KLULA	P20386 kluyveromyc
21	92	3.6	535	1 COX1_HANWI	P48868 hansenula w
22	92	3.6	898	1 ACOC_CUCMA	P49608 cucurbita m
23	92	3.6	4367	1 DYHC_NEUCR	P45443 neurospora
24	91.5	3.6	1259	1 LIN1_HUMAN	P08547 homo sapien
25	91	3.5	1032	1 MT18_YEAST	P40469 saccharomyc
26	91	3.5	1054	1 POL_SIVMK	P05897 simlan immu
27	90	3.5	573	1 DPOL_MOUSE	Q9qxe2 mus musculus
28	90	3.5	948	1 RPOD_FODAN	Q01521 podospora a
29	90	3.5	1125	1 TIE2_BOVIN	Q06807 bos taurus
30	90	3.5	1300	1 POL2_MOUSE	P11369 mus musculus
31	89.5	3.5	412	1 SYI2_BACSU	P25151 bacillus su
32	89	3.5	678	1 UVRB_PASMU	P57844 pasteurella
33	88.5	3.4	286	1 PARB_CHLPI	Q9z7m0 chlamydia p

34	88.5	3.4	341	1 YF43_MYCTU	Q10783 mycobacteri
35	88.5	3.4	536	1 DIT1_YEAST	P21623 saccharomyc
36	88.5	3.4	859	1 ABR_HUMAN	Q12979 homo sapien
37	88.5	3.4	949	1 RCSC_ECOLI	P14376 escherichia
38	88	3.4	984	1 SECA_AQUAE	O67718 aquifex aeo
39	88	3.4	1106	1 CIC2_RABIT	P13806 oryctolagus
40	88	3.4	3137	1 CA3B_CHICK	P15989 gallus gall
41	87.5	3.4	476	1 ATPB_GALSU	Q08807 gaidieria s
42	87.5	3.4	534	1 COX1_SACDO	P98001 saccharomyc
43	87.5	3.4	557	1 COX1_NEUCR	P03945 neurospora
44	87.5	3.4	577	1 SYMM_CANAL	O74634 candida alb
45	87.5	3.4	627	1 MUTL_BACSU	P49850 bacillus su

ALIGNMENTS

RESULT 1
T3MH_HAEIN
ID T3MH_HAEIN STANDARD; PRT; 629 AA.
AC P71366;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative type III restriction-modification system HindVIP enzyme mod
DE (EC 2.1.1.72) (HindVIP methyltransferase) (M.HindVIP).
GN HI1056.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd".
RL Science 269:496-512(1995).
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC -!- SUBUNIT: CONTAINS TWO DIFFERENT SUBUNITS: RES AND MOD. MOD IS
CC A HOMOTETRAMER (BY SIMILARITY).
CC -!- SIMILARITY: WITH OTHER TYPE III MOD PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32786; AAC22721.1; -
CC REBASE; 3701; M.HindORF1056P.
CC TIGR; HI1056; -
CC InterPro; IPR002295; D2IN6_mtfrase.
CC InterPro; IPR001091; N4_Mtase.
CC InterPro; IPR002052; N6_Mtase.
CC InterPro; IPR002941; N6_N4_Mtase.
CC Pfam; PF01555; N6_N4_Mtase; 1.
CC PRINTS; PR00506; D21N6MTFRASE.
CC PRINTS; PR00508; S21N4MTFRASE.
CC PROSITE; PS00092; N6_MTASE; 1.
CC Hypothetical protein; Transferase; Methyltransferase; DNA-binding;

KW

KW Restriction system; Complete proteome.
SQ SEQUENCE 629 AA; 71845 MW; 93ADAD909DBA1B84 CRC64;

Query Match 3.9%; Score 101.5; DB 1; Length 629;
Best Local Similarity 19.9%; Pred. No. 3.3;
Matches 54; Conservative 50; Mismatches 98; Indels 69; Gaps 9;

QY 33 VNFNLSLQLDPSIIHGWDFLSKRLVTLARGLSPAFRFGKRTDFLOFQNLRPAK 92
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 198 IEQAYLKILMDIDIFDRNFLNTIVTKTIGGVSGSE-----GSKLDSTEFINVFKNR 252
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 93 SRGGCPDY-----YLKNYEDD-----IVRSVDALDKQCKCTIAQHPDWMLE 134
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 253 ERLFNPVYQTEVNEFIKNYEDSGSKMYQTQVLIDLGEKILLEKDGFKYYHYPN--- 308
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 135 LQREKAQHMLVLLKPFQSFNTYSNILLPEPN--VRTMHGRAVNGSQLGDYIOLKSLLQ 192
||| : : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 309 -----AQMTSIVKFSDQNLSKEIITYEYSHKVYRTTNAQSSTRSKIIEDLYSIKNGIV 362
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 193 PIRIYSRASLYGNIGRPKNVIALL-----DGFPMKVGSTV-----D 230
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 363 SIE-----YIPQKGKAGNLIEFYFNASNKDMFMELSLMIKEKNKYFYLQKVNTLWD 415
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 231 AVTWQC-----YIDGRVKVMDFLTRLLD 256
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 416 DIQYNLNLEGGYIDFKNGKKPEALLRIID 446

RESULT 2
CYAA_YEAST STANDARD; PRT; 2026 AA.

ID CYAA_YEAST
AC P08678;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Adenylate cyclase (EC 4.6.1.1) (ATP pyrophosphate-lyase) (Adenylyl
cyclase).
GN CRY1 OR CDC35 OR HSR1 OR SRA4 OR YJL005W OR J1401.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86079531; PubMed=2934138;
RA Kataoka T., Brock D., Wigler M.;
RT "DNA sequence and characterization of the S. cerevisiae gene encoding
adenylate cyclase."
FT Cell 43:493-505(1985).
RN [2]
RP SEQUENCE FROM N.A.
RA To Van D., Perea J., Jacq C.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RA de Haan M., Smits P.H.M., Grivell L.A.;
RL submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE OF 1042-2026 FROM N.A.
RX MEDLINE=88165073; PubMed=3327602;
RA Masson P., Lenzen G., Jacquemin J.M., Danchin A.;
RT "Yeast adenylate cyclase catalytic domain is carboxy terminal.";
RN Curr. Genet. 10:343-352(1986).
RN [5]
RP MUTAGENESIS OF THR-1651.
RX MEDLINE=91122042; PubMed=1991451;
RA Feger G., de Vendittis E., Vitelli A., Masturzo P., Zahn R.,
RA Verrotti A.C., Kavounis C., Fal G.P., Fasano O.;
RT "Identification of regulatory residues of the yeast adenylyl
cyclase."
RL EMBO J. 10:349-359(1991).

```
FT CONFLICT 1566 1566 A -> S (IN REF. 1).
FT CONFLICT 1735 1735 V -> G (IN REF. 1).
FT CONFLICT 1956 1956 I -> V (IN REF. 3).
FT CONFLICT 1996 1996 C -> F (IN REF. 1).
FT CONFLICT 2009 2026 NVDELLOWVKNAKDLST -> MLTLNFYKWLRTQRYOLE
FT FCS (IN REF. 3).
SQ SEQUENCE 2026 AA; 227832 MW; BEAB7419AB6989D0 CRC64;

Query Match 3.9%; Score 101; DB 1; Length 2026;
Best Local Similarity 20.1%; Pred. No. 18;
Matches 120; Conservative 80; Mismatches 201; Indels 196; Gaps 30;

QY 12 AGLKE-KTLLDVTKNPV---RTVNENFLSLQDPSI-----I 47
||| ||| ||| ||| : : : ||| ||| :
Db 903 AGFVELKNQLLDLSNFKMFHYPEVINYCTNLLQIDLSYNKIQSLPQSTKYLVKLAKMNL 962
||| ||| ||| ||| : : : ||| ||| :
QY 48 HDGWDELSSKRLVLARGLSPAFRFGCKRTDFLQFQNL----- 87
||| ||| ||| ||| : : : ||| ||| :
Db 963 SHNKLNFGDLSEMDLRTLNLRNIRISSIKTKNASNQLNQLFTDNRISNFDLTPKLRLAL 1022

QY 88 ---RNPASRGPGDYLYLKNVEDDITVRSVDALDKGCKIAOHDPVMLE---LQREKA 140
||| ||| ||| ||| : : : ||| ||| :
Db 1023 ETQENPITS--FKDFPKN-----WTSLLNK---AQLSSIPGELLTKLSFLEKLEL 1071
||| ||| ||| ||| : : : ||| ||| :
QY 141 AQHHLVLKKEQFSNTYSNLIILTEPNN-----YRT--MHGRAVNGSQLGKDY 184
||| ||| ||| ||| : : : ||| ||| :
Db 1072 NQNNLTRLPQETSKLTKLVFLSVARNKLEYIPPELSQLSKSLRTLDLHSNNIRDFVDGMEN 1131

QY 185 IQLKSLQPIRYSRASL-----YCPNIGRPRKNVIALLDGFMKVAGSTVDVAVTQ 235
||| ||| ||| ||| : : : ||| ||| :
Db 1132 LETSLNTISNFAFGSSLENSYHNMSYSGSKLSK-----SLM--FFIAADNQFDDAMWP 1183

QY 236 --HCYIDGRVAVKMFDFLTRLDLTSDQIRKIQKVNTYTPCKKIWLEGVVTTSAGTNN 293
||| ||| ||| ||| : : : ||| ||| :
Db 1184 LFNCFN---LKVNLNYSNFSV---SHMKLESTIELYSGNKL-----TTLSDGT-- 1229

QY 294 LSDSYAAGFLW--LNTGLMLANGIDV-----VIRHSFFDGHYNHLVDQNFNLPDY-W 344
||| ||| ||| ||| : : : ||| ||| :
Db 1230 -----VLKWSLTKTLMNSQMLSLPAELSNLSQSVDFVGANQKYNISYHYDWNW 1282

QY 345 LS---LLYKRLGIP-----KVLAVHAGLQKRPGRVI 375
||| ||| ||| ||| : : : ||| ||| :
Db 1283 RNKKELYNLFNGNRRRFEIKSPISHIDADLSLTVLPOLKVLGLMDVTLNTTKVPDENY 1342

QY 376 RDKLRIYAHCTHHNH-----NYV-----RGSITLFTIINLHRSKK----- 411
||| ||| ||| ||| : : : ||| ||| :
Db 1343 NFRLRTASINGMYGVADTLGQRDYVSSRDVTFERFGNDECLCLHDSKNQADYG 1402

QY 412 ----IKLAGFLRK-LVHQYLLQPYGOE-----GLKSKSVQLNGO---PLVMVDDG 454
||| ||| ||| ||| : : : ||| ||| :
Db 1403 HNISRIVRDIYDKILIRQ--LERYGDETDDNIKTALRFSFLQLNKEINGMLNSVDNG 1457

RESULT 3
GIDA_TREPA
ID GIDA_TREPA STANDARD; PRT; 630 AA.
AC O83084;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glucose inhibited division protein A.
GN GIDA OR TP0044.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
```

```
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT *Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.;
RL Science 281:375-388(1998).
CC -!- FUNCTION: NOT KNOWN.
CC -!- SIMILARITY: BELONGS TO THE GIDA FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001189; AAC65038.1; -.
DR TIGR; TP0044; -.
DR InterPro; IPR002218; GIDA.
DR Pfam; PF01134; GIDA; 1.
DR ProDom; PD003738; GIDA; 1.
DR PROSITE; PS01280; GIDA_1; 1.
DR PROSITE; PS01281; GIDA_2; 1.
DR Complete proteome.
SQ SEQUENCE 630 AA; 70175 MW; 1B52C6F02C1EC275 CRC64;

Query Match 3.9%; Score 100.5; DB 1; Length 630;
Best Local Similarity 21.3%; Pred. No. 3.9;
Matches 77; Conservative 52; Mismatches 129; Indels 103; Gaps 20;

QY 177 GSOLGK--DYIOLKSLLOPIRIYSRASLYCPNIGRPRKNVIALLDGFMKVAGSTVDVAVTW 234
||| ||| ||| ||| : : : ||| ||| :
Db 68 GGEWGFADACMIQ-----YRLNKR--GPAVQAPRIQADKFL--YAQKVYLTLETOH 118

QY 235 QHCYIDGRVAVKMFDFLTRLDLTSDQIRKIQKVNTYTPCKKIWLEGVVTTSAGTNNL 294
||| ||| ||| ||| : : : ||| ||| :
Db 119 LHLQDQ---TVVDVVCSTNTDAGYVAYGAHAVVT--ARGRRISARAVVLTG----- 166

QY 295 SDSYAAGFLW-----NVLGMLANGIDVTVIRHSFFDHG-----YHVLVDQNF 337
||| ||| ||| ||| : : : ||| ||| :
Db 167 --TFMGRIYVIGYEAPGRLGEHAAEGALRKGFGQMGRLKGTGPAPVLRSKSD--- 221

QY 338 NLPDYWLSLYKR---LIGPKVLAVHAGLQKRPGRVIRDKLRIYAHCTHHNHNY 393
||| ||| ||| ||| : : : ||| ||| :
Db 222 -----LSVMEKQEDAINRPFSA-HV-----EINRPHADCYINYTNERTHQL 263

QY 394 VRGSITLFTIINLHRS---RKKIKLAGT-----LRDKLVHQYLLQPYGOEGLKS 438
||| ||| ||| ||| : : : ||| ||| :
Db 264 IRE-----NFHRSPPFSGRKAVGTRVYCPSTEDKVRKFPDRIRHQLYIEP---EGLDT 313

QY 439 KSVQLNGQPLVMVDD-----GTLPELK-----PRPLRAGRTLVIPVYMGF-FVKNVNA 487
||| ||| ||| ||| : : : ||| ||| :
Db 314 EELYINGLSCLPEDIQDEMIRTPGMEIRAVITRPAYAVDYAVLFPVLQIDLTQTRVSG 373

QY 488 L 488
Db 374 L 374

RESULT 4
SYW_BUCAI
ID SYW_BUCAI STANDARD; PRT; 335 AA.
AC P57602;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Tryptophan--tRNA ligase)
DE (TRRS).
GN TRPS OR B0536.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum)
```

[illegible]

88	RNPAKSRGGPGPDYLYLKNYEDDVRSDVALQKGCKIAHQHPDMVLELQREKAAQ----	142
Qy	: : :	
Ddb	: : :	
84	RGKVLGKG-----FAKCY-----MTDLTNKKVYAANKIIPHSRVAKPHOREKDIEILH	134
Qy	: : :	
Ddb	: : :	
143	-----MHLVLCLKEQFSNTYSNLILTEPNMYRTMHGRAVNSQLGKYIQLKSLQP--IRI	196
Qy	: : :	
Ddb	: : :	
135	RILHHKHVOGYHYFEDKENIVILLEYCSRSM-----AHILK-----ARKVLTPEVRY	184
Qy	: : :	
Ddb	: : :	
197	YSRASLYXGNIGRPKNRNVIALLDGFMKVGAGSTVDATWQHQCIDGRVVVMDFLKTRLDD	256
Qy	: : :	
Ddb	: : :	
185	YLROIVSGLKYLHEQE-----ILHRDLKLGNFFIN-----EAMELKVKYDGDFGLAARLE	231
Qy	: : :	
Ddb	: : :	
257	TLSDOIIRKIQQKVNVTYP-----GKKTLWEGVV--TTSAG-----CTNNLSDSY--	298
Qy	: : :	
Ddb	: : :	
232	PLEHRRRTICGTPNLYSFEVLNKQGHCESDIWALGCVMYTMLLRGPFFETTNLKETYRC	291
Qy	: : :	
Ddb	: : :	
299	--AAGFLWLNTL-----GMLAQ-----GIDWTRHSHFPDHGY--NHLWDQNENPL	340
Qy	: : :	
Ddb	: : :	
292	IREARYTMPSSLLAPAKHLIASMLSKNPBDPRLSDIIDIRHDFFOGTFPTORLSSSCCHTV	351
Qy	: : :	
Ddb	: : :	
341	PBYWTSTAIVKPRIGPKVIHVHVAGLORKPR-----PGRVIRDKLRITAYACHTCNHNHNHNVYRG	396
Qy	: : :	
Ddb	: : :	

QY		188	KSLLOPIRISRASL-YGPNIGRPKNVTAL-LDGFMKVAGSTVDAVTWOHCYID--GR	242
	:	:	: :: :	:
Db		31	KTLLMTINLVAFPCNPISGGSAGKIVVREIDALDGGEMGNRIDTKTYIQMKMLNTGKP	90
	:	:	: :: :	:
QY		243	VVKVM-----DFLKTRLLDTLSQ-----IRKI-----QKVNYTYT-PGKKIWL	280
	:	:	: :: :	:
Db		91	AVRALRAQADKDEYADSMMKNTYSQENLTROGMVEELIDEEKKKVIKGTSTGTKYA	150
	:	:	: :: :	:
QY		281	EGVYTT-----SAGGTNMLSRYANGFLWNLTLGLMLANOGIDVVIRHSF	324
	:	:	: :: :	:
Db		151	KAVIIITGTALRGTEITIGELKYSSSGNNLSL-----SIGUADN-----LRIGT	193
	:	:	: :: :	:
QY		325	FDHGYNHVLVDQNENPLDPYMWLSLLYKRLLGPVKYLAVHVHAGLRKPRPG-----	372
	:	:	: :: :	:

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

[illegible]

Query Match	3.6%;	Score 93;	DB 1;	Length 488;
Best Local Similarity	21.2%;	Pred. No. 10;		
Matches	82;	Conservative	49;	Mismatches 141; Indels 114; Gaps 21;

Qy	21	LLDVSTKN---PV--RVVNENFSLQLODPSI--THDGWL-----DFLSSKRLVTLARGLSPA	70
Db	106	LLDVTSVNEETPIVVKRPINERIELEEDPRFEDQAKLEISEDKLGSGFFGEVGYL--L	163
Qy	71	FLRFGGKRTDFLOQLNRNPAKSRGGPGDYILKNEYEDDIRSDVALDKQCKCIAQHPD	130
Db	164	SHRSTNTETDYL--QKLSVAVKQSNPTQENQEKMIED-----ETKLMCAIGRNP	212
Qy	131	VM-----LELQREAAQMHLV-----LLK--EFSNTYSNLLITEPNNYRTMHGRAVN	176
Db	213	ILAIIGAVTANGSARNLLIIEFVECGDLLKFLFEKKSIFKDELVEYKNGY-----	263
Qy	177	GSQGLKDYIQLKSLLOPIRIYSRASLYCPNIGRPKNVI--ALLDGMFKVAGSTVDVAWTQ	235
Db	264	-----LLPKSIRKTYMFNEN-----EDDVEESLDL-----CTSDLLSFS	300
Qy	236	HCYIDG-----RVKVMDFLKLRLDLSQIRKIQRVWT	271
Db	301	YQIAREMEYLASIPCVHRDLALRWLNKNTIRIADFGLRKYQV--DGYRITKGVGT	358
Qy	272	YTPGGKIWLGVVTTTSAGGTNNLSD--SYAAGFLWNLNTGLMLANQGI---DVVIRHSFFD	326
Db	359	PMPAR--WMAPEVMRE--KCTCEKSDVMSYGSVLYEMESLIGELPYSNVNSDV-----	407
Qy	327	HGYNLHVQDNFNPLPDYWLSLLYRKL	352
Db	408	--FEHVQGNOLNPMPOYCHPKMYDRM	431

RESULT_14
HSG_SYN3
ID
PASC_SYN3
STANDARD;
E7351.
DT
30-MAY-2000 (Rel. 39, Created)
DT
30-MAY-2000 (Rel. 39, Last sequence update)
DT
16-OCT-2001 (Rel. 40, Last annotation update)
DE
Glycogen phosphorylase (EC 2.4.1.1).

```

GN GLCP OR SL11356.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxId=1148;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugliura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RA "Sequence analysis of the genome of the unicellular cyanobacterium
FT Synechocystis sp. strain PCC6803. II. Sequence determination of the
FT entire genome and assignment of potential protein-coding regions.";
RL Ent Res. 3:109-136(1996).
CC -1- FUNCTION: PHOSPHORYLASE IS AN IMPORTANT ALLOSTERIC ENZYME IN
CC CARBOHYDRATE METABOLISM. ENZYMES FROM DIFFERENT SOURCES DIFFER IN
CC THEIR REGULATORY MECHANISMS AND IN THEIR NATURAL SUBSTRATES.
CC HOWEVER, ALL KNOWN PHOSPHORYLASES SHARE CATALYTIC AND STRUCTURAL
CC PROPERTIES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: {(1,4)-alpha-D-glucosyl}(N) + phosphate =
CC {(1,4)-alpha-D-glucosyl}(N-1) + alpha-D-glucose 1-phosphate.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE GLYCOGEN PHOSPHORYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D0907; BAA17551.1; -.
DR HSP; D06738; 11CP.
DR InterPro; IPR000811; Phosphorylase.
DR Pfam; PF00343; phosphorylase; 1.
DR PROSITE; PS00102; PHOSPHORYLASE; 1.
DR Transferase; Glycosyltransferase; Carbohydrate metabolism;
KW Glycogen metabolism; Pyridoxal phosphate; Complete proteome.
FT BINDING 679 679 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 849 AA; 96648 MW; 657D3CE5B3C23A3D CRC64;

```

Query Match 3.6%; Score 93; DB 1; Length 849;
Best Local Similarity 19.8%; Pred. No. 22;
Matches 94: Conservative 51; Mismatches 130; Indels 200; Gaps 25;

```

QY 32 TVNENFSLQLDPSSIIHDGWLDFLSSKRRLVTLARGLSPAFLRFGCGKRTDELQFQNLRNPA 91
Db 465 TILKDYEYL-----WPEKFSNK-----TNGVTP-----RRMWLSNPRLSNLI 502
QY 92 KSRGGPGDYLYLKNYEDDIR-----SDVALDKQGCKIAQ-----HP 129
Db 503 SSRICDG---WIKNL-DELKOLEPADLAGFRQDWCKVKREKVQDLARYIHTRTLVLVNP 558
QY 130 DYMLELQ-----REKAAQHMLVLLKEOFSNTYSNLIITEPNNRYTMHGRVANGSOLG 181
Db 559 DSLFDQVQVRHEHYKRQHMLHLVTHLYLQIKNN--PNLDVT-PTTF-----IYGRAA 609
QY 182 KDYIQLSLLOP IRIYSRASLYGPNIGRPKRKNVIALLDGFMKVGASTVDVATVWQHICYIDG 241
Db 610 PGYETAKLIILINSADVWNNDPTIG-DRLXVFIPDYNVK-----FGQ 653
QY 242 RVVKVMDELFKTRLLDTLSQIRKIQQVVNTYPGKKIWLEGVVTTSAGCTNNLSDSYAAG 301
Db 654 RVPAPAD-----LSEQIS-----TAGKE-----ASCTGNMKFMSMGA 685
QY 302 FLWLNTLGLMANQGDIV-----318
Db 686 L-----TIGTLDGCANTEIREEVGAENFFULGTTPTEVEKTLASGYQPWEYNNANLKV V 741

```

```
QY 319 -VIRHSFEDHGNHVLVDQNFNLPDYWLWLLKRLIGPKVLAVHVGLOKPRGVRVD 377
Db 742 DLINGFFSHGDTAL-----FRPLMD-----SLGQD-----PYLVFAD 775
QY 378 KLRIYAHCTNH-----HHNNVVRGSIPTLFIINLHRSRKKIKLACTLRDKLKHQY 426
Db 776 -FOAYVDQNVQGEAYKDOENWAR-----MALLNVARWGK-----FSSDRTIREY 819

RESULT 15
SNK_MOUSE STANDARD; PRT; 682 AA.
AC P53351;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Serine/threonine-protein kinase SNK (EC 2.7.1.-) (Serum inducible
DE kinase).
DE SNK.
GN SNK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92375085; PubMed=1508211;
RA Simmons D.L., Neel B.G., Stevens R., Evett G., Erikson R.L.;
RT "Identification of an early-growth-response gene encoding a novel
RT putative protein kinase.";
RL Mol. Cell. Biol. 12:4164-4169(1992).
CC -!- FUNCTION: MAY PLAY A ROLE IN THE DIVISION OF AT LEAST SOME CELL
CC TYPES, SUCH AS FIBROBLASTS, AND COULD FUNCTION IN EMBRYOGENESIS,
CC WOUND HEALING OR NEOPLASIA.
CC -!- TISSUE SPECIFICITY: BRAIN, LUNG AND HEART.
CC -!- INDUCTION: BY SERUM AND PHORBOL ESTER.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC CDC5/POLO SUBFAMILY.
CC -!- SIMILARITY: CONTAINS 2 POLO BOX DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M96163; -; NOT_ANNOTATED_CDS.
DR HSP; P00518; 1PHK.
DR MGD; MGI:1099790; Snk.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000959; POLO_box.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00659; POLO_box; 2.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS50078; POLO_BOX; 2.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00101; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding; Repeat.
FT DOMAIN 54 59 POLY-HIS.
FT DOMAIN 79 331 PROTEIN KINASE.
FT NP_BIND 85 93 ATP (BY SIMILARITY).
FT BINDING 108 108 ATP (BY SIMILARITY).
FT ACT_SITE 202 202 BY SIMILARITY.
FT DOMAIN 507 570 POLO BOX 1.
FT DOMAIN 603 674 POLO BOX 2.
SQ SEQUENCE 682 AA; 77811 MW; 586DEABFD7208A9D CRC64;
```

Query Match 3.68; Score 92.5; DB 1; Length 682;
Best Local Similarity 20.9%; Pred. No. 18;

```
Matches 89; Conservative 58; Mismatches 167; Indels 111; Gaps 19;
QY 88 RNPASRGPGPDYLYLKNYEDDIVRSDVALDKQKGIQAHPDVMLELQREKAAQ-----142
Db 81 RCKVLGKGG-----FAKCYE-----MTDLTNKVVYAAKIIPHRSVAKPHQREKIDKEIELH 131
QY 143 -----MHLVLLKEQFSNTYSNLITPNNYRTMHGRAVNGSOLGKDYIOLKSLQIP-IRI 196
Db 132 RLLHHKHVVQFYHYFEDKENIYILLEYCSRSM-----AHILK-----ARKVLTEPEVRY 181
QY 197 YSRASIYGNIGRPNKRNVTALLDGMKVGASTVDATWQHCVYIDGRVVKVMDFLKTRILD 256
Db 182 YLRQIVSGLKYLHEQE-----ILHRLDKLGNFFIN-----EAMELVGDFGLAARLE 228
QY 257 TLSDQIRKIQKVNTYTP-----GKKIWLEGV--TTSAG-----GTNNLSDSY-- 298
Db 229 PLEHRRRTICGTPNVLSPVLNKGHCESDIWALGCVMYTMLLGRPPETTNLRETYRC 288
QY 299 ---AAGFLMLNTL-----GMLANQ-----GIDVVIHRSFFDHGY--NHLVDQNFNPL 340
Db 289 IREARYTMPSSLLAPAKHLLIASMLSKNPNDRPSLDDIIRHDFFLQGFTPDRLSSSCCHTV 348
QY 341 PDYWLSSLYKRLIGPKVLAVHVGLOKPRGVRVDKLRIVAHCTNHHNNHNVVRGSIYL 400
Db 349 PDFHLSSPAKNFF-KKAAAALFEGGKKDKAR-----YNDTHNKVSK 387
QY 401 FTINLHRSRKKIKLAGTLRDKLVHVVLLQPY-----GOEGLKSKSVQLNGOPLVMV 451
Db 388 EDEDIYKLRHDLKKVSIQOQSKHRADEEPQPPPTVARSGETSAVENK--QQIGDAIRMI 445
QY 452 DDGTL 456
Db 446 VRGTL 450

Search completed: July 30, 2002, 08:31:30
Job time: 977 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:30:43 ; Search time 123.94 Seconds
(without alignments)
688.127 Million cell updates/sec

Title: US-09-836-461-2_COPY_42_534
Perfect score: 2577
Sequence: 1 GDRRLPVDRAAGLKEKTLI.....PVTMGFFVKNVNALACRYR 493

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query	Length	DB	ID	Description
1	2577	100.0	534	4	Q9HB38	Q9hb38 homo sapien
2	2538	98.5	592	4	Q9HB37	Q9hb37 homo sapien
3	2270	88.1	480	4	Q9HB39	Q9hb39 homo sapien
4	953	37.0	536	11	Q9QZF8	Q9qzf8 rattus norv
5	927.5	36.0	545	4	Q9UL39	Q9ul39 homo sapien
6	926.5	36.0	543	4	Q9Y251	Q9y251 homo sapien
7	915.5	35.5	545	6	Q9MY0	Q9my0 bos taurus
8	835.5	32.4	523	13	Q90YK5	Q90yk5 gallus gall
9	302	11.7	521	10	Q9SDA1	Q9sda1 arabidopsis
10	302	11.7	543	10	Q9FF10	Q9ff10 arabidopsis
11	261.5	10.1	516	10	Q9FLK8	Q9flk8 arabidopsis
12	256.5	10.0	527	10	Q9LRC8	Q9lrc8 scutellaria
13	246	9.5	536	10	Q9FZP1	Q9fzp1 arabidopsis
14	171	6.6	935	5	Q9VE79	Q9ve79 drosophila
15	155	6.0	190	10	O82604	O82604 arabidopsis
16	113	4.4	4533	5	Q9BIX3	Q9bix3 tetrahymena

17	111	4.3	1260	5	O15639	O15639 dictyosteli
18	110	4.3	174	10	Q9ATW5	Q9atw5 zea mays (m
19	106.5	4.1	885	5	O61232	O61232 lymnaea sta
20	106	4.1	768	3	Q9P8N9	Q9p8n9 cladospriu
21	105	4.1	1020	5	O25415	O25415 leishmania
22	101.5	3.9	1234	5	Q9Y1H7	Q9y1h7 dictyosteli
23	100.5	3.9	582	17	Q9HRZ5	Q9hrz5 halobacteri
24	99.5	3.9	455	3	Q12229	Q12229 saccharomyc
25	99.5	3.9	518	5	Q968Y8	Q968y8 caenorhabdi
26	99.5	3.9	535	5	Q968V7	Q968v7 caenorhabdi
27	99.5	3.9	1272	16	Q98PR8	Q98pr8 mycoplasma
28	99	3.8	1713	9	Q94MA1	Q94mal lactococcus
29	99	3.8	2403	5	O15881	O15881 trypanosoma
30	99	3.8	4578	13	O42181	O42181 fugu rubrip
31	98.5	3.8	1214	12	Q9INJ1	Q9inj1 banna virus
32	98	3.8	491	2	O08457	O08457 clostridium
33	98	3.8	907	10	Q9SE90	Q9se90 brassica ol
34	98	3.8	1090	12	Q91E95	Q91e95 human rotav
35	98	3.8	1281	11	O88915	O88915 mus musculu
36	97.5	3.8	515	5	Q24109	Q24109 drosophila
37	97.5	3.8	553	5	Q9VEQ5	Q9veq5 drosophila
38	97.5	3.8	1167	16	Q9A205	Q9a205 streptococc
39	97	3.8	838	10	Q9FG37	Q9fg37 arabidopsis
40	96.5	3.7	1429	2	Q9LAP7	Q9lap7 alteromonas
41	96	3.7	763	10	Q9ZUL5	Q9zul5 arabidopsis
42	96	3.7	898	10	Q9SIB9	Q9sib9 arabidopsis
43	96	3.7	1219	11	Q9QYM3	Q9qym3 mus musculu
44	95.5	3.7	406	2	O57513	O57513 coxiella bu
45	95.5	3.7	472	16	Q98QF0	Q98qf0 mycoplasma

ALIGNMENTS

RESULT 1

Q9HB38	PRELIMINARY;	PRT;	534 AA.
ID	Q9HB38;		
AC	Q9HB38;		
DT	01-MAR-2001 (Tremblrel. 16, Created)		
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)		
DT	01-MAR-2001 (Tremblrel. 16, Last annotation update)		
DE	HEPARANASE-LIKE PROTEIN HPA2B.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20483645; PubMed=11027606;		
RA	McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,		
RA	Hircock M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;		
RT	"Cloning and Expression Profiling of Hpa2, a Novel Mammalian		
RT	Heparanase Family Member."		
RL	Biochem. Biophys. Res. Commun. 276:1170-1177(2000).		
DR	EMBL; AF282886; AAG23422.1; -.		
SQ	SEQUENCE 534 AA; 60663 MW; C3DE5E900CB338C4 CRC64;		

Query Match 100.0%; Score 2577; DB 4; Length 534;
Best Local Similarity 100.0%; Pred. No. 1.2e-198;
Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHGWLDLSSKRL	60
DB	42	GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHGWLDLSSKRL	101
QY	61	VTLAGLSLPAFLFGKRTDFLQFNLRNPAKSRGPGPDYLLKNYEDDIVRSDVALDKQ	120
DB	102	VTLAGLSLPAFLFGKRTDFLQFNLRNPAKSRGPGPDYLLKNYEDDIVRSDVALDKQ	161
QY	121	KCKTAQHPDVMLEQREKAAQMHVLVLLKEQFSNTYSNLITPEPNRYTMHGRAVNGSOL	180
DB	162	KCKTAQHPDVMLEQREKAAQMHVLVLLKEQFSNTYSNLITPEPNRYTMHGRAVNGSOL	221

Qy	181	GKDYIQLKSLLOPIRIYRSASLYGPNIGRPRKNVIALLDGFMKVAGSTVDVATWQHCHYID	240
Db	222	GKDYIQLKSLLOPIRIYRSASLYGPNIGRPRKNVIALLDGFMKVAGSTVDVATWQHCHYID	281
Qy	241	GRVVKVMDFLKTRLLDLSDOIIRKIQKVVNTYTPGKKIWLGVVTTSGAGTNNLSDSYAA	300
Db	282	GRVVKVMDFLKTRLLDLSDOIIRKIQKVVNTYTPGKKIWLGVVTTSGAGTNNLSDSYAA	341
Qy	301	GFLWLTNLGLANOGIDVVIIRHSFFDHGYNHLVDQNFNPLDPYWLSSLYKRLIGPKVLAV	360
Db	342	GFLWLTNLGLANOGIDVVIIRHSFFDHGYNHLVDQNFNPLDPYWLSSLYKRLIGPKVLAV	401
Qy	361	HVAGLQRPGRGVRTRDKLRIYAHCTNNHNNHNVYRGSTITLFIINLHRSRKKIKLAGTLRD	420
Db	402	HVAGLQRPGRGVRTRDKLRIYAHCTNNHNNHNVYRGSTITLFIINLHRSRKKIKLAGTLRD	461
Qy	421	KLVHGYLLQPYGQGLSKSVOLNGQPLVMVDDGTLPELKPRLRAGRTLVIPTVMGFF	480
Db	462	KLVHGYLLQPYGQGLSKSVOLNGQPLVMVDDGTLPELKPRLRAGRTLVIPTVMGFF	521
Qy	481	VVKVNNALACRYR 493	
Db	522	VVKVNNALACRYR 534	
RESULT	2		
Q9HB37			
ID	Q9HB37	PRELIMINARY;	PRT; 592 AA.
AC	Q9HB37;		
DT	01-MAR-2001 (Tremblrel. 16, Created)		
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)		
DT	01-MAR-2001 (Tremblrel. 16, Last annotation update)		
DE	HEPARANASE-LIKE PROTEIN HPA2C.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20483645; PubMed=11027606;		
RA	McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,		
RA	Hircock M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;		
RT	"Cloning and Expression Profiling of Hpa2, a Novel Mammalian		
RT	Heparanase Family Member."		
RL	Biochem. Biophys. Res. Commun. 276:1170-1177(2000).		
DR	EMBL; AF282887; AAG23423.1; -.		
SQ	SEQUENCE 592 AA; 66580 MW; 95C384AD9A74258E CRC64;		
Query Match	98.5%;	Score 2538;	DB 4; Length 592;
Best Local Similarity	89.5%;	Pred. No. 1.9e-195;	
Matches 493;	Conservative 0;	Mismatches 0;	Indels 58; Gaps 1;
Qy	1	GDRRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNNFSLQLDPSIIHDGWLDFLSSKRL	60
Db	42	GDRRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNNFSLQLDPSIIHDGWLDFLSSKRL	101
Qy	61	VTIARGLSPAFIRFGGKRTDFLQFNLRNPAKSRGGGPDYILKNYEDDIVRSDVALDKQ	120
Db	102	VTIARGLSPAFIRFGGKRTDFLQFNLRNPAKSRGGGPDYILKNYEDDIVRSDVALDKQ	161
Qy	121	KGCKIAQHPDVMLEQREKAQHLLVLLKEQFSNTYSNLITL-----	162
Db	162	KGCKIAQHPDVMLEQREKAQHLLVLLKEQFSNTYSNLITL-----	221
Qy	163	-----EPNNYRTMHGRAVNSQLGK	192
Db	222	IFALNLRNPNNSWSSSALSLLKYSASKYNIWSWELGNEPNPNYRTMHGRAVNSQLGK	281
Qy	183	DYIQLKSLLOPIRIYRSASLYGPNIGRPRKNVIALLDGFMKVAGSTVDVATWQHCHYIDGR	242
Db	282	DYIQLKSLLOPIRIYRSASLYGPNIGRPRKNVIALLDGFMKVAGSTVDVATWQHCHYIDGR	341

Qy	243	VVKVMDFLKTRLLDLSDOIIRKIQKVVNTYTPGKKIWLGVVTTSGAGTNNLSDSYAAGF	302
Db	342	VVKVMDFLKTRLLDLSDOIIRKIQKVVNTYTPGKKIWLGVVTTSGAGTNNLSDSYAAGF	401
Qy	303	LWLTNLGLANOGIDVVIIRHSFFDHGYNHLVDQNFNPLDPYWLSSLYKRLIGPKVLAVH	362
Db	402	LWLTNLGLANOGIDVVIIRHSFFDHGYNHLVDQNFNPLDPYWLSSLYKRLIGPKVLAVH	461
Qy	363	AGLQRPGRGVRTRDKLRIYAHCTNNHNNHNVYRGSTITLFIINLHRSRKKIKLAGTLRDKL	422
Db	462	AGLQRPGRGVRTRDKLRIYAHCTNNHNNHNVYRGSTITLFIINLHRSRKKIKLAGTLRDKL	521
Qy	423	VHGYLLQPYGQGLSKSVOLNGQPLVMVDDGTLPELKPRLRAGRTLVIPTVMGFFV	482
Db	522	VHGYLLQPYGQGLSKSVOLNGQPLVMVDDGTLPELKPRLRAGRTLVIPTVMGFFV	581
Qy	483	KVNNALACRYR 493	
Db	582	KVNNALACRYR 592	
RESULT	3		
Q9HB39			
ID	Q9HB39	PRELIMINARY;	PRT; 480 AA.
AC	Q9HB39;		
DT	01-MAR-2001 (Tremblrel. 16, Created)		
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)		
DT	01-MAR-2001 (Tremblrel. 16, Last annotation update)		
DE	HEPARANASE-LIKE PROTEIN HPA2A.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20483645; PubMed=11027606;		
RA	McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,		
RA	Hircock M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;		
RT	"Cloning and Expression Profiling of Hpa2, a Novel Mammalian		
RT	Heparanase Family Member."		
RL	Biochem. Biophys. Res. Commun. 276:1170-1177(2000).		
DR	EMBL; AF282887; AAG23421.1; -.		
SQ	SEQUENCE 480 AA; 53900 MW; F75F89F67AC1FF83 CRC64;		
Query Match	88.1%;	Score 2270;	DB 4; Length 480;
Best Local Similarity	89.0%;	Pred. No. 4.8e-174;	
Matches 439;	Conservative 0;	Mismatches 0;	Indels 54; Gaps 1;
Qy	1	GDRRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNNFSLQLDPSIIHDGWLDFLSSKRL	60
Db	42	GDRRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNNFSLQLDPSIIHDGWLDFLSSKRL	101
Qy	61	VTIARGLSPAFIRFGGKRTDFLQFNLRNPAKSRGGGPDYILKNYEDDIVRSDVALDKQ	120
Db	102	VTIARGLSPAFIRFGGKRTDFLQFNLRNPAKSRGGGPDYILKNYEDDIVRSDVALDKQ	149
Qy	121	KGCKIAQHPDVMLEQREKAQHLLVLLKEQFSNTYSNLITLPEPNYRTMHGRAVNSQL	180
Db	150	-----EPNNYRTMHGRAVNSQL	167
Qy	181	GKDYIQLKSLLOPIRIYRSASLYGPNIGRPRKNVIALLDGFMKVAGSTVDVATWQHCHYID	240
Db	168	GKDYIQLKSLLOPIRIYRSASLYGPNIGRPRKNVIALLDGFMKVAGSTVDVATWQHCHYID	227
Qy	241	GRVVKVMDFLKTRLLDLSDOIIRKIQKVVNTYTPGKKIWLGVVTTSGAGTNNLSDSYAA	300
Db	228	GRVVKVMDFLKTRLLDLSDOIIRKIQKVVNTYTPGKKIWLGVVTTSGAGTNNLSDSYAA	287
Qy	301	GFLWLTNLGLANOGIDVVIIRHSFFDHGYNHLVDQNFNPLDPYWLSSLYKRLIGPKVLAV	360
Db	288	GFLWLTNLGLANOGIDVVIIRHSFFDHGYNHLVDQNFNPLDPYWLSSLYKRLIGPKVLAV	347

QY 361 HVAGLQKPRGVRIRDKLRIYAHCTNNHHNNHNVYRGSTLFIINLHRSRKKIKLAGTLRD 420
|||||
Db 348 HVAGLQKPRGVRIRDKLRIYAHCTNNHHNNHNVYRGSTLFIINLHRSRKKIKLAGTLRD 407
QY 421 KLVHQLLQPYGOEGLSKSVOLNGQPLVMVDDGTLPELKPRLRAGRTLVIPTVTMGFF 480
|||||
Db 408 KLVHQLLQPYGOEGLSKSVOLNGQPLVMVDDGTLPELKPRLRAGRTLVIPTVTMGFF 467
QY 481 VVKNNALACRYR 493
|||||
Db 468 VVKNNALACRYR 480
RESULT 4
Q90ZF8 PRELIMINARY; PRT; 536 AA.
AC Q90ZF8;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE HEPARANASE.
GN HEP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Podyma K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;
RT "Heparanase from parathyroid cell line."
RL Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF184967; AAF04563.1; .
SQ SEQUENCE 536 AA; 60569 MW; 6208B1FD9EE28421 CRC64;

Query Match 37.0%; Score 953; DB 11; Length 536;
Best Local Similarity 39.3%; Pred. No. 3.9e-68;
Matches 216; Conservative 83; Mismatches 165; Indels 86; Gaps 10;
QY 1 GDRRLPVDRAAG-LKEKTLILLDVSTKNPVRTVNENFLSLQDPSIIHD-GWLDPLFSK 58
Db 12 GLRLALTQGTAGTAPTKDQVLEFYTKRLFGSVSPSLTIDASLATDPRFLTLGSP 71
QY 59 RLRLTARGSLPAFLRGGKRTDQLQNLNPAKSRGGPGDPYLYKNYEDDVRSDVALD 118
Db 72 RURALARGSLPAYLRFGGTKTDLIF----DPNKEPTSEERSYQSQDNDICGSE---- 123
QY 119 KQKCKIAQHPDYMELQREKAAQHLVLLKEQF-----SNTYS----- 157
Db 124 -----RVSADVLRKQWEPFQ-ELLLREQYREFKNSTYSRSDVDMLYSFAKCSRL 175
QY 158 -----NLIL-----TEPNYRTMHGRAVNGSOL 180
Db 176 DLIFGLNALLRTPDLRWSSNAQLLLNYCSSKGYNISWELGNPNFSFKKADIFINGSQLGEDEFQLHKLRLK-S 235
QY 181 GKDYIOLKSLQPIRIYSRASLYGNIPGRKNVIALLDGFMKVGSTVDVATWQHCHYD 240
Db 236 GEDFVELHKLQK-SAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLYN 294
QY 241 GRVVKVMDFLKRLDLSQIRKTKQVYNTYTPGKKIWLGVVTTTSAGTNNLSDSYAA 300
Db 295 GRVATKEDFLSSDVLDTFLSQIKLVTKEMTPGKKVWLGETSSAYGGGAPLLSNTFAA 354
QY 301 GFLWLNTLGMANQGDIVYRHSFDDHGYNHLVDQNFNPLPDYWSLLYKRLIGPKVLAV 360
Db 355 GFWMLDKGLSQAQLEGVVNRQVFFAGNYHLVDENFPLPDYWSLLYKRLIGPKVLAV 414
QY 361 HVAGLQKPRGVRIRDKLRIYAHCTNNHHNNHNVYRGSTLFIINLHRSRKKIKLAGTLRD 420
Db 415 RVKQPD-----RSKLRYVILHCTNNVHPRYREGDLYLVNLHNVTKHLKLPMPMS 465
QY 421 KLVHQLLQPYGOEGLSKSVOLNGQPLVMVDDGTLPELKPRLRAGRTLVIPTVTMGFF 480

Db 466 RVDYLLKPFSGDGLLSKSVOLNGQTLKXVDEQTLPALTEKPLPAGSSLSYPAFSYGF 525
QY 481 VVKNNALAC 490
Db 526 VVIRNAKIAAC 535
RESULT 5
Q9UL39 PRELIMINARY; PRT; 545 AA.
AC Q9UL39;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE HEPARANASE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=20229546; PubMed=10764835;
RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;
RT "Heparanase expression in invasive trophoblasts and acute vascular damage."
RL Glycobiology 10:467-475(2000).
DR EMBL; AF084467; AAD54516.1; .
SQ SEQUENCE 545 AA; 61418 MW; 67B80ACD73C5A9A1 CRC64;
Query Match 36.0%; Score 927.5; DB 4; Length 545;
Best Local Similarity 38.9%; Pred. No. 4.5e-66;
Matches 208; Conservative 78; Mismatches 162; Indels 87; Gaps 10;
QY 16 EKTLLLDVSTKNPVRTVNENFLSLQDPSIIHD-GWLDPLFSKRLVTLARGSLPAFLRF 74
Db 37 QQDVLDLDEFTQERHLVSPSFLSVTIDANLATDPRFLILGSPKRLTLARGSLPAYLRF 96
QY 75 GSKRTDQLQNLNPAKSRGGPGDPYLYKNYEDDVRSDVALDQKQCKIAQ-HPDVML 133
Db 97 GGTKTDLIF----DPKKESTFEERSYQSQVQNDI-----CKYGSIPPDVEE 140
QY 134 ELQREKAAQHLVLLKEQF-----SNTYS-----NLIL----- 161
Db 141 KLRLEWQYQEQ-LRHHYQKFKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADL 199
QY 162 -----TEPNYRTMHGRAVNGSOLGKDYIOLKSLQPIR 195
Db 200 QWNSSNAQLLLDYCSSKGYNISWELGNPNFSFKKADIFINGSQLGEDEFQLHKLRLK-S 258
QY 196 IYSRASLYGNIPGRKNVIALLDGFMKVGSTVDVATWQHCHYDGRVVKVMDFLKTRL 255
Db 259 TFKNAKLYGPDVQPRKTKAKMLKSLKAGGEVIDSVTHHYLYNGRTATREDFLAPDVL 318
QY 256 DTLSQDQIRKTKQVYNTYTPGKKIWLGVVTTTSAGTNNLSDSYAAAGFLWLNTLGMANQ 315
Db 319 DIFISSVQVQVQVVESTRPGKKVWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMG 378
QY 316 IDVIRHSFDDHGYNHLVDQNFNPLPDYWSLLYKRLIGPKVLAVHAGLQKPRGVR 375
Db 379 IEVVMRQVFFGAGNYHLVDENFPLPDYWSLLYKRLIGPKVLAVHAGLQKPRGVR 431
QY 376 RDKLRYAHCTNNHHNNHNVYRGSTLFIINLHRSRKKIKLAGTLRDLKLVHQLLQPYGOE 435
Db 432 --KLRYVILHCTNNVHPRYREGDLYLVNLHNVTKHLKLPMPMS 489
QY 436 LKSKSVOLNGQPLVMVDDGTLPELKPRLRAGRTLVIPTVTMGFFVVKNNALAC 490
Db 490 LLSKSVOLNGTLTKWDDQTLPLMEKPLRPGSSGLPFAFSYFFVIRNAKVAAC 544

```

RESULT 6
Q9Y251 PRELIMINARY; PRT; 543 AA.
ID Q9Y251
AC Q9Y251
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE HEPARANASE.
GN HPA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=99321249; PubMed=10395326;
RA Hulet M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,
RA Parish C.R.;
RT "Cloning of mammalian heparanase, an important enzyme in tumor
RT invasion and metastasis."
RL Nat. Med. 5:803-809(1999).
RN [2]
SEQUENCE FROM N.A.
RA Vlodavsky I., Friedman Y., Elkin M., Aingorn H., Atzmon R.,
RA Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I.,
RA Spector L., Pecker I.;
RT "Mammalian heparanase: a novel gene involved in tumor progression and
RT metastasis."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE FROM N.A.
RX MEDLINE=99377052; PubMed=10446189;
RA Toyoshima M., Nakajima M.;
RA "Human heparanase. Purification, characterization, cloning, and
RT expression."
RL J. Biol. Chem. 274:24153-24160(1999).
RN [4]
SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=9935379; PubMed=10405343;
RA Kussie P.H., Hulmes J.D., Ludwig D., Patel S., Navarro E.C.,
RA Seddon A.P., Giorgio N.A., Bohlen P.;
RT "Cloning and Functional Expression of a Human Heparanase Gene."
RL Biochem. Biophys. Res. Commun. 261:183-187(1999).
DR EMBL; AF165154; AAD43379.1; -
DR EMBL; AF144325; AAD41342.1; -
DR EMBL; AF155510; AAD54941.1; -
DR EMBL; AF152376; AAD45669.1; -
SQ SEQUENCE 543 AA; 61176 MW; AD262EC267334AB2 CRC64;

Query Match 36.0%; Score 926.5; DB 4; Length 543;
Best Local Similarity 38.5%; Pred. No. 5.4e-66;
Matches 212; Conservative 79; Mismatches 170; Indels 89; Gaps 11;

QY 1 GDRPLPVDRAAGLKEKTLILLDVTSTKPNVTVNENFLSLQDPSIIHD-GWLDLFLSKSR 59
DB 22 GLPSGALPRA-QAQDVLDFTQEPHLVSPFSLVTDANLATDPRFLILGSPK 79
QY 60 LVTLAGLSPALRFGGKRTDPLQFQNLNRPAPKSRGGPGDPYLLKNYEDDIDVRSDDALDK 119
DB 80 LRTLARGLSPALRFGGKRTDPLF-----DPKKESTFEERSYQSQVNQDI----- 126
QY 120 QKGCIAQ-HPDVMLELQREKAAQMHVLVLLKEQF-----SNTYS----- 157
DB 127 ---CKYGSIPDVEEKLRLWPYQQL--LLREHYQKFKNSTYSRSSVDVLYTFANCSSL 182
QY 158 -----NLIL-----TEPNNYRTMHGRAVNSQL 180
DB 183 DLIFGLNALLRTADLQWSSNAQLLDYCSSKGYNISWELGNEPNSFLKADIFINGSOL 242
QY 181 GKDYIQLKSLQPIRIYSRASLYGFGNIGRPKNVTLDDGFMKVGASTVDATVQHCYID 240

RESULT 7
Q9MY0 PRELIMINARY; PRT; 545 AA.
ID Q9MY0
AC Q9MY0;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HEPARANASE.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Kizaki K., Nakano H., Takahashi T., Imai K., Hashizume K.;
RT "Expression of Heparanase mRNA in Bovine Placenta During Gestation."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF281160; AAF87301.2; -
SQ SEQUENCE 545 AA; 61076 MW; FAC4BDFD85B933 CRC64;

Query Match 35.5%; Score 915.5; DB 6; Length 545;
Best Local Similarity 37.6%; Pred. No. 4.2e-65;
Matches 204; Conservative 77; Mismatches 169; Indels 93; Gaps 9;

QY 7 PVDRAAGLKEKTLILLDVTSTKPNVTVNENFLSLQDPSIIHD-GWLDLFLSKRLVTLAR 65
DB 36 PADDAAE-----LEFFTERPLHLVSPAFSLFTIDANLATDPRFTFLGSSKRLTLAR 87
QY 66 GLSPAFLRGGKRTDPLQFQNLNRPAPKSRGGPGDPYLLKNYEDDIDVRSVALDKQKCKI 125
DB 88 GLAPALRFGGKRTDPLFDPKPEAFEE-----RSWLSQSQNDI-----CKS 131
QY 126 AQHPDVMLELQREKAAQMHVLVLLKEQF-----SNTYS-----NLIL--- 161
DB 132 GSIPSDVEEKLRLWPYQQLRQYQKFTNSTYSRSSVDMLYTFASCGLNLIYFGV 191
QY 162 -----TEPNNYRTMHGRAVNSQLGKDYIOL 187
DB 192 ALLRTTDMHWDSSNAQLLDYCSSKGYNISWELGNEPNSFORKAGIFINGRQCGEDFIEF 251
QY 188 KSLIOPRIYSRASLYGFGNIGRPKNVTLDDGFMKVGASTVDATVQHCYIDGRVVVKM 247
DB 252 RKLGLK-SAFNAKLYGPDIGQPRNTVKKLSFLKAGGEVIDSVTWHYVNGRIATKE 310
QY 248 DFKTRLLDLSQIIRKIQKVVNTYTPGKKIWLVEGVVTSAGGTNNLSDSYAAGFLWNT 307

```

```
Db 311 DFLNPDILDTFISSVQKTLRIVEKIRPLKVKWVLGETSSAFGGGAPFLSNTPFAAGFWMLDK 370
Qy 308 LGMLANQIDVYIRHSFHDGYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHVAGLQR 367
Db 371 LGLSARMGIEVVMQVLFAGAGNYHLVDGNFELPDYWLSSLLFKVLGNKVLMSYKGPD- 429
Qy 368 KPRGRVTRDKLRIYAHCTNHHNNHNVYRGSTITLFIINLHRSRKKIKLAGTLRDKLVHQYL 427
Db 430 -----RSFRVYLHCTNTKHPRYKEGDTLYALNLHNVTKHELPHHLFNKQVDKYL 481
Qy 428 LPQYGOGLKSKSVOLNGOPLVMVDGTLPELKPRLRAGRTLVIPTVMTGFFVYKVNVA 487
Db 482 IKPSTGDLKSKSVOLNGOILKXVDEQTLPALTEPLHPGSSLGMPFPSPSYGFFVIRNAKV 541
Qy 488 LAC 490
Db 542 AAC 544

RESULT 8
ID Q90YK5 PRELIMINARY; PRT; 523 AA.
AC Q90YK5;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HEPARANASE.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=11387326;
RA Goldsmith O., Zcharia E., Aingorn H., Guatta-Rangini Z., Atzmon R.,
RA Michal I., Pecker I., Mitrani E., Vlodaysky I.,
RT "Expression Pattern and Secretion of Human and Chicken Heparanase Are
RT Determined by Their Signal Peptide Sequence.";
RL J. Biol. Chem. 276:29178-29187(2001).
DR EMBL; AY037007; AAK92648.1; -.
SQ SEQUENCE 523 AA; 58386 MW; 8EB0B7B18C9BFB81 CRC64;

Query Match 32.4%; Score 835.5; DB 13; Length 523;
Best Local Similarity 36.6%; Pred. No. 1.1e-58;
Matches 192; Conservative 74; Mismatches 180; Indels 79; Gaps 8;

Qy 22 LDVSTKNPVRTVNFNLSLQIDPSIIHD-GWLDFLSSKRLVTLARGLSAPFLRFGKRTD 80
Db 21 LQGLREPIGAVSPAFSLTLDASLARDPRFVALLRHPKLHTLASGLSPGLRFGGTSTD 80
Qy 81 FLQFNLRNPAKSRGGPGPDYLYKNVEDIVRSDVALD-----KQ 120
Db 81 FLIF-----NPNKD-----STWEEKVLSEFOAKDCEAMPFVVPKLLLTQWPIQ 126
Qy 121 KGCKIAOHP-----DVM-----LELQREKAAQMH 145
Db 127 EXLLLAHSSKKHKNNTTITRSLDILHTFASSSGFLVFNALLRAGLQWSSNAKOL 186
Qy 146 VLLKQFSNTYNNLITPEPNNTYTHMGRAVNGSQGDKYIQKLSLLOPTRYSRASLYGP 205
Db 187 LGYCAQRSYNIWEELNENFNSRKRKSGICIDGFLGRDFVHLRQLLSQHPLYRHAELYL 246
Qy 206 NIGRPRKNVIALLDGFMKVAGSTVDATVWOHCYIDGRVVKVMDLTKRLTLDLSOIRKI 265
Db 247 DVGQPKKTHQLRSEFMKSGGKAIDSVTWHHYVNGRSATREDFLSPEDLSFATAIHQV 306
Qy 266 QKVNTYTPGKKIWLLEGVVTTSAGTNNLSDSYAAGFLMLNTGLMLANOGIDVIRHSFF 325
Db 307 LGIVATVPKKVWLGETGSAYGGAPQLSNITYVAGFMWLDKLGLAARSGIDVVMRQVSF 366
Qy 326 DHGYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHVAGLQRKPRGRVTRDKLRIYAH 385
```

```
Db 367 GAGSYHLVDAGPKPLPDYWLSSLYKRLVGRVLRQASVEQADAR-RP-----RYVLHC 417
Qy 386 TTHNNHNVRGSTITFIINLHRSRKKIKLAGTLRDKLVHQYLLOPVGOBGLKSKSVOLNG 445
Db 418 TNPRHPKREGDVTFLALNLSNVTQSLQPKQLWSKSDVOYLLLPHGKDSILSREVQLNG 477
Qy 446 QPLVMVDGDTLPKLPRLRAGRTLVIPTVMTGFFVYKVNVA 490
Db 478 RLQMVDDDTLPALHEMALAPGSTLGLPAFSYGFVIRNAKAIAC 522

RESULT 9
ID Q9SDA1 PRELIMINARY; PRT; 521 AA.
AC Q9SDA1;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE HYPOTHETICAL 57.8 KDA PROTEIN.
GN F13G24.30.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Van Der Schueren J., Chuang Y.J., Voet M., Robben J.,
RA Voickaert G., Bancroft I., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL13421; CAB62595.1; -.
DR InterPro; IPR001254; Trypsin.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 521 AA; 57831 MW; 07D8664A4B305CC2 CRC64;

Query Match 11.7%; Score 302; DB 10; Length 521;
Best Local Similarity 24.8%; Pred. No. 7.9e-16;
Matches 139; Conservative 73; Mismatches 183; Indels 166; Gaps 27;

Qy 34 NENFLSLQID--PSIIHD-----GWLDFLS---SKRLVTLA-RGLSPAFLRFGGKR 78
Db 22 DENFVCATLDWMP---HDKCNYDQCWPYSSVINMDLTRPLLTAKAKPKLRIRIGSSL 78
Qy 79 TDFLQFO--NLNPA----KSRGGPGPDYLYKNVEDIVRSDVALDKQKCKIAQHDPVM 132
Db 79 QDQYIDVGNLKTPCRPFOKMNSG-----LFGFSKGCLHMKRWD-- 117
Qy 133 LELQREKAAQMHV-----LLKEQFSNTYSNLITPEPNNTYTHMGRAVN---- 176
Db 118 -ELNSLTATGAVVTPLGNALRGHKLRCGAWGGAMDHINTQDFLNTYKSGVYDSWEF 176
Qy 177 GSOL-----GKDYIQLKSLLOPTRYSRASLYGNIGRPRKNVIALLDGF- 221
Db 177 GNELSGSGVGSVSAELYKGLDVLVLDVIN--KVYKNSWLHKPILVAP-----GGFY 226
Qy 222 -----MKVAG-STVDATVWOHCYIDGR-----VKVMDFLKTRLLDLSQIRKIQ 266
Db 227 EQQWYTKLLEISGSPVVDVVT-HHYNLGSNDPALVKKIMD---PSYLSQVSKTFKDVN 282
Qy 267 KVVNTYTPGKKIWLLEGVVTTSAG-----GTNNLSDSYAAGFLMLNTGLMLANOGIDVIRH 322
Db 283 QTIQEHGP---WASPWGESGGAYNSGGRHVSDDTFIDSFYLDQLGMSARHNTKYVCRO 338
Qy 323 SFEDHCYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHVAGLQRKPRGRVTRDKLRIY 382
Db 339 TLVGGFYGLLEKGTFFPNPDYYSALLWHRLMGKGVLAQVDTG-----PPQLRVY 387
```

Db 361 TLVGGFVLEKGTFFVPNDYYSALLWHLRMGKGLAVQTDG-----PPQLRVY 409

Qy 383 AHCTNHHNNHYRGSTITLFIINLH-----RSRKKIKLAGTLRDKLV 423

Db 410 AHCSKG-----RAGVTLLILNLSNQSDFTVSVSNGINNVVLNAESRKKKSLDITLKRPF 463

Qy 424 -----HOYLQPYGQEG-LKSKSVOLNGQPLVMVDDGTLPELKPRLRAGRT 469

Db 464 WIGSKASDGYLNREYHLTP--ENGVLRSKTMVLNGSLKPTATGDIPLSLEPVLRSVNSP 521

Qy 470 LVIPPTMGFFVVKVNNALAC 490

Db 522 LNVPLSMSFIVLPNFDASAC 542

RESULT 11

Q9FLK8 PRELIMINARY; PRT; 516 AA.

AC Q9FLK8;

DT 01-MAR-2001 (TremBLrel. 16, Created)

DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)

DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)

DE SIMILARITY TO HEPARANASE.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RX MEDLINE=98290546; PubMed=9628582;

RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,

RA Tabata S.;

RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV. Sequence features of the regions of 1,456,315 bp covered by nineteen physically assigned P1 and TAC clones."

RL DNA Res. 5:41-54 (1998).

DR EMBL; AB010073; BAB08480.1; -

DR InterPro; IPR001917; AminoTransf_2.

DR PROSITE; PS00599; AA_TRANSFER_CLASS_2; UNKNOWN_1.

SQ SEQUENCE 516 AA; 57414 MW; 6943E6D3D89FB93C CRC64;

Query Match 10.1%; Score 261.5; DB 10; Length 516;

Best Local Similarity 21.1%; Pred. No. 1.4e-12;

Matches 121; Conservative 94; Mismatches 201; Indels 157; Gaps 22;

Qy 16 EKTLLLDVSTKNPVTVNENFLSLQLD-----PSIIHOGWLDLSSKR 59

Db 2 ERTLLVDGSR--TAETDENFICATLDWPEKCNKYDQCPWGYASLINLN---LASPL 55

Qy 60 LVTLAGRLSPAFRFGKRTDFLQF--NLNRPASRGGGPDYLYKNVEDDIVRSDVAL 117

Db 56 LAKAIQAFRTLRIIRIGSLQDQVIYDVGDLKTPCTQ-----FKTDDGL-----F 100

Qy 118 DKQGGCKIAQHPDVMLELQREKAA---QMHLVLLKEQFSNT-----YSNLILTEPNYR 168

Db 101 GFSEGLYMKRWDEVHNFFNATGAIYVFGNLALHGRNKLNGTAWGGDWDHTNQDFMNYT 160

Qy 169 TMHGRAVN-----GSQ-----GKDYIQKLSLLQPIRYSRASLYGPNIGRP 210

Db 161 VSKGYALDSWEFGNELSGSGIWAASVSVELYKDLVLKNVKNVKNRSRT-----210

Qy 211 RKNVIALLDGF-----MKVAGSTVDVATWQHICYDQ-----RVVKVMDFLKTRLL 255

Db 211 -KPLVAPGQFFEEQYSELLRSLSGPGLDVLTHHLYNLGPGNDPKLVNKILD---PNYL 266

Qy 256 DTLSQDIKTKQVNTVTGCKIWLGVVTTTSAGTNN-----LSDSYAAGFLWNLTLGM 310

Db 267 SGISELFANVQTIQEHGFWAAAW-----GEAGGAFNSGGQVSETFINSFWLQDLGI 321

Qy 311 LANOGIDVIRHFFDGHYNHLVDQNFNPLDYLWLSLLYKELIGPKVLAVHVGALQKRP 370

Qy 383 AHCTNHHNNHYRGSTITLFIINLH-----RSRKKIKLAGTLRDKLV 423

Db 388 AHCSKG-----RAGVTLLILNLSNQSDFTVSVSNGINNVVLNAESRKKKSLDITLKRPF 441

Qy 424 -----HOYLQPYGQEG-LKSKSVOLNGQPLVMVDDGTLPELKPRLRAGRT 469

Db 442 WIGSKASDGYLNREYHLTP--ENGVLRSKTMVLNGSKLPTATGDIPLSLEPVLRSVNSP 499

Qy 470 LVIPPTMGFFVVKVNNALAC 490

Db 500 LNVPLSMSFIVLPNFDASAC 520

RESULT 10

Q9FF10 PRELIMINARY; PRT; 543 AA.

AC Q9FF10;

DT 01-MAR-2001 (TremBLrel. 16, Created)

DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)

DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)

DE SIMILARITY TO HEPARANASE.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RX MEDLINE=97471969; PubMed=9330910;

RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,

RA Miyajima N., Tabata S.;

RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence features of the 1.6 Mb regions covered by twenty physically assigned P1 clones."

RL DNA Res. 4:215-230 (1997).

DR EMBL; AB005249; BAB09947.1; -

DR InterPro; IPR001254; TrypSIN.

DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.

SQ SEQUENCE 543 AA; 60250 MW; 0FA2248948282FF6 CRC64;

Query Match 11.7%; Score 302; DB 10; Length 543;

Best Local Similarity 24.8%; Pred. No. 8.4e-16;

Matches 139; Conservative 73; Mismatches 183; Indels 166; Gaps 27;

Qy 34 NENFLSLQLD--PSIIHD-----GWLDFLS---SKRLVTLA-RGLSPAFLRFGKR 78

Db 44 DENEVCATLDWPP---HDKCNKYDQCPWGYSSVINMDLTRPLLTAKAKFKPLRIRIGGSL 100

Qy 79 TDFLOFQ--NLNRPAA---KSRGGPGDYLYKNVEDDIVRSDVALDKQCKIAQHPDVM 132

Db 101 QDQVIYDVGNDUKTPCRPFQKNSG-----LFGSKGLHMKRWD-- 139

Qy 133 LELOREKAAQMHV-----LLKEQFSNTYSNLILTEPNNYRTMHGRAVN-----176

Db 140 -ELNSFLTATCAVVTFGNLALRGRHLRGKAGGAWDHIINTQDFLNTVTSKGYVDSWEF 198

Qy 177 GSQ-----GKDYIQKLSLLQPIRYSRASLYGPNIGRPKNVIALLDGF- 221

Db 199 GNELSGSGVGASVSAELYKDLVLKOVIN--KVKNKSWLHKPLIVAP-----CGFY 248

Qy 222 -----MKVAG-STVDAVTWQHICYDGR-----VVKVMDFLKTRLLDPLSQIRKI 266

Db 249 EQQWTYKLEISGPSVVDVVT-HHHYNLGSGNDPALVKKIMD---PSYLSQVSKTFKDVN 304

Qy 267 KVVNTYTGCKIWLGVVTTTSAG-----GTNNLSDSYAAGFLWNLTLGMLANOGIDVIRH 322

Db 305 QTIQEHGP-----WASPMWGESGAYNSGGRHVSDFIDSFYLDOLGMSARHNTKVYCRQ 360

Qy 323 SFFDGHYNHLVDQNFNPLDYLWLSLLYKELIGPKVLAVHVGALQKRPGRVIRDKLRIY 382

Db 322 SSKHNTKCYCRQALVGGFVGLLEKETFPVNPDPYYSALLWHRLMGKILGVQTTA----- 375
QY 371 PGRVIRDKLRIYAHCTNHHNNHNVGRSGITLFIINLHRSKIKLA-----SRKK 411
Db 376 -----SEYLRAYVHCSCR-----RAGIYILLINLSKHHTTFVAVSNGVKKVVLQAESMR 424
QY 412 IKLAGFLRDKLV-----HOYLLQPYGQEG-LKSKSVQVNGOPLVMVDDGTLP 457
Db 425 KGFLETIKSKSVWVGKASDGYLNREYHLSR--KDGDLRSKIMLLNGKPLVPTATGDIP 482
QY 458 ELKPRPLRAGRRLVIPPVTMGFFVKKVNNALAC 490
Db 483 KLEPVRHGKSPVYINPLSISFVLPTFDAPAC 515
RESULT 12
Q9LRC8 PRELIMINARY; PRT; 527 AA.
AC Q9LRC8;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE BETA-GLUCURONIDASE.
GN SGUS.
OS Scutellaria baicalensis.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Lamiales; Lamiaceae; Scutellaria.
OX NCBI_TaxID=65409;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20418130; PubMed=10858442;
RA Sasaki K., Taura F., Shoyama Y., Morimoto S.;
RT "Molecular Characterization of a Novel beta-Glucuronidase from
RT Scutellaria baicalensis Georgi.";
RL J. Biol. Chem. 275:27466-27472(2000).
DR EMBL; AB040072; BAA97804.1; -;
DR InterPro; IPR001179; FKBP_PPase.
DR PROSITE; PS00453; FKBP_PPase_1; UNKNOWN_1.
SQ SEQUENCE 527 AA; 58772 MW; A5DE7C423F2A1E2B CRC64;

Query Match 10.0%; Score 256.5; DB 10; Length 527;
Best Local Similarity 24.2%; Pred. No. 3.6e-12;
Matches 129; Conservative 81; Mismatches 207; Indels 115; Gaps 27;
QY 15 KEKTLILLDVSTKNPVTNENFLSLQD---PSIIHDG---W-----LDFLSKRLV 61
Db 26 EETIVKIE---ENPVAQDENTVCATLDLWPPTKCYNGCPWKGKSSFLNLD-LNNNIIR 81
QY 62 TLARGLSPAFRLRGGKRTDFLOFQNLNRPAAKSRGGPGPDYLYKNYEDDIVRSVDALDKQ 121
Db 82 NAVKEFAPLKLREGGTLQDLRYVQT-----SRDECDSTFYNN-----TNLIDFSH 128
QY 122 GKIAQHPDV-----LELOREKAAQMHVLKKEQ-----FSN 154
Db 129 ACLSLDRWDEINQFILETGEAVFGLNLRGKTVETIKQKQYLGETTAVGDEWDSYN 188
QY 155 TYSNLILTEPNRYRTHMGRAVNGSQLG-----KDYI-QLKSLLQPIR-IYRSASL 202
Db 189 SKFLIYSKKGKYGKRGWTL--GNEIGGHTLFTGVSPEDYANDAKKLHELKVLKFIYDQGT 247
QY 203 YGPNIGRPRKNVIALLDGFMKVGAGSVDA-VTWHQCYIDGRVVVKVDFELKTRLLD-TLSD 260
Db 248 M-PLIATP--GAIFDELWTEFTDRPELHVATHMYNLSGGD--DALKDVLTTASFDD 302
QY 261 QIRK-----IQKVVNTYTPGKK--IWLEGVVTTSAGTNNLSDSYAAGFLWLNTLGLMAN 313
Db 303 EATKSMYEGKQIVN--RPTKAVAMIGEAGGAFNSGQDGISNTFFNGFWYLNWLYSAL 360
QY 314 QGIDVIVIRISFFDHGYNHLVDQNFNLPDYWLSLLKRLIGPKVLAVHVAGLQRPGR 373

Db 361 LOTKTEFCROTLLTGGNYGLLQGTGYIPNPDPYYSALLWHRLMGSKVLKTEIVGTK----- 413
QY 374 VIRDKLRIYAHCTNHHNNHNVGRSGITLFIINLHRSKIKLA-----GTLRDKLVHGYL 427
Db 414 -----NVIYAHCAKKN-----GITMLVLN-HGSSSVKISLDPSKYSKRE-----EYH 458
QY 428 LQPYGQEGLKSKSVQVNGOPLVMVDDGTLPKLPRLRAGRRLVIPPVTMGF 479
Db 459 LTPV-NNNLQSLRVKLGELLHLDPSGVIPALNPVEKDNQKQLEVAPYSFMF 509
RESULT 13
Q9FZP1 PRELIMINARY; PRT; 536 AA.
AC Q9FZP1;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE SIMILARITY TO HEPARANASE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB028613; BAB10787.1; -;
SQ SEQUENCE 536 AA; 59654 MW; 24202B8E82F3DB0E CRC64;
Query Match 9.5%; Score 246; DB 10; Length 536;
Best Local Similarity 21.4%; Pred. No. 2.6e-11;
Matches 119; Conservative 86; Mismatches 195; Indels 156; Gaps 25;
QY 30 VRTVNEFLSLQD---PSIIHDG---WDLFLSKRL-----VTLARGLSPAFRLRFGG 76
Db 41 VGTIDEDFICATLDWPPKCDYGCSCSW-DHASILNLDLTIVVPVPSKNTAFAPLKRIGG 99
QY 77 KRTDFLOFQNLNRPAAKSRGGPGPDY-----YLNKTYEDDIVRSVDALDKQKCKIAQHPD 130
Db 100 TLQDIVIYET-----PDSKQPCLPFTKN-----SSILFGYTQGCPLMRWD 140
QY 131 VMLELOREKAAQMHVL-----LKE-----QFSNTYSNLILTEPNRYR----- 168
Db 141 ELNAFFRKTKTVIFGLNALSGRSIKSNGEAGAWNYTNAESFIRFTAENNYTIDGWELG 200
QY 169 -----TMHGRAVNGSQLGKDYIOLKSILQPIRIYSRAS---LYGPNIGRPRKNVIALLD 219
Db 201 NELCGSGVGARVAGANQAYDITINLRIVN--RVYKNVSPMLVIGPG----- 245
QY 220 GFMKV-----AGSTVDVAVTWHQCY-----IDGRVVKVMDFLKTRLDLTLSQIRKI 265
Db 246 GFPEVDWETFLNKAENSLNATT-RHYIDLGPGVDEHLIE--KILNPSYLDQEAQKFSRL 302
QY 266 QKVVNTYTPGKKIWLGVVTTTSAG-----GTNNLSDSYAAGFLWLNTLGLMANOGIDVIR 321
Db 303 KNTIKNSSTKAVAW----VGESGGAYNSGRNLVSNAFVSFWYLDQGLMASLYDTKYCR 358
QY 322 HSFEDHCYNHLVDQNFNPLPDYWLSLLKRLIGPKVLAVHVAGLQRPGRVIRKLR 381
Db 359 QSLIGNGYGLLNTNTPNDPYYSALIMQLMGRKALFTTFSGTK-----KIRS 407
QY 382 YAHCTNHHNNHNVGRSGITLFIINLHRSR-----KKIKLAGTLRDKLVHGYLQPY----- 431
Db 408 YTHCAQSK-----GITVLLMNLNDNTTVAKVELNNSF--SLRTHKHKMSYKRASSQL 459
QY 432 --GOEG-----LKSXSVQVNGOPLVMVDDGTLPKLPRLRAGRRLVIPP 474

Db 460 FGGPNCVIOREYHHTAKDGNLHNSQTMLLNGNALQVNSMGDLPPIEPHINSTEPIITAP 519
Qy 475 VTMGFEVKNVNALAC 490
Db 520 YSIVFVHMRNVVPAC 535

RESULT 14
Q9VE79 PRELIMINARY; PRT; 935 AA.
AC Q9VE79;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE CG14309 PROTEIN.
GN CG14309.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132.
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazie R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003721; AAF55548.1;
DR Flybase; FBgn0038611; CG14309.
SQ SEQUENCE 935 AA; 106883 MW; 10E46B721CE5DB3F CRC64;

Query Match 6.6%; Score 171; DB 5; Length 935;
Best Local Similarity 20.9%; Pred. No. 6.3e-05;
Matches 106; Conservative 76; Mismatches 163; Indels 162; Gaps 26;

Qy 18 TLILLDVSTKNPVRTVNFENFLSLQLDPSLIHGWLDLSSRLVTLARGLSPAFLR----- 73

Db 19 TLVLQHVN--RPYNDVNEKFVSFAVRPEDLYDA-LDGKRRKAVTNLANLLGDHAIKAVNW 75
Qy 74 -----FGGKRTDFLOFQNLNRPASRGPGPDYLYKNYEDDI--VRSDVALDKQKCK 124
Db 76 TMIIPVPYAPDDWDMSMTLKILNTS-----YMGVITDCIWLQGTDTGTSRAK--- 122
Qy 125 IAQHPDVMLELOREKAAQMHVLVLLKEQFSNTYSNLIITEPNNYRTMHGRAVNGSQLG--K 182
Db 123 -----DIYQELRLTKL-----MTDTFKPVV-----DDWRLMGADISAGSSADETK 162
Qy 183 DYIQLKSL-----QPIRYSRASLYGPNIGRPKNVIALLDGFMKVAGSYDVAVTWQ 235
Db 163 RYVDMSKDLNTAFGTQTPANMLPKSSL-----GS----- 191
Qy 236 HCYIDGRVVKVMDFLKTRLLDLSDOIIRKIOKVNVNTYTPGKKIWLGVVVTTSAGCTNNLS 295
Db 192 YLYDSDPALRTLQOORVPLWUTLPEE-RSSQRLV-----GDETDA----- 231
Qy 296 DSYAAGFLWNLTLGLMNOGIDVIRHSFFDHGYNHLVD---QNFNPLPDYWLSLYKRL 352
Db 232 -----LRWQTMGDAAASGDFVIFKRM-----NLVDFERPNS-----LYVTALEKVV 274
Qy 353 IGPKVLAHVHAGLQKRPGRVIRDKLRIYAHCTNHHNHYVRSITLFIINLHRSKKI 412
Db 275 MGRSVFPA-----RPLNAFAPSNKLYTHCA-----NAVSGLAFMVNVTEEQPTTI 320
Qy 413 ---KLAGTLRDKLVHOLYLPQYGOGLKSKSVOLGQPLVMVDDGTL-PELK- ---RPLR 465
Db 321 TVKSTSSSSSSEIWIQVLTGHDQR-----VOLNNVRLHL--NTTLRPLIKPIDPTKPLQ 372
Qy 466 AGRTLVPVPTMGFFVVKNNVNALACRY 492
Db 373 ----LITPSMAVSFWVLPDVLNLEHCQF 395

RESULT 15
082604 PRELIMINARY; PRT; 190 AA.
ID 082604;
AC 082604;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE T2L5.6 PROTEIN.
GN T2L5.6.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Geisel C., Smith A., Le T.;
RT "The sequence of A. thaliana T2L5";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Washu;
RT "The A. thaliana Genome Sequencing Project";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Waterston R.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF096371; AAC62794.1; --
SQ SEQUENCE 190 AA; 21212 MW; B5E356E5526007A0 CRC64;

Query Match 6.0%; Score 155; DB 10; Length 190;
Best Local Similarity 24.9%; Pred. No. 0.00011;


```
Matches 49; Conservative 32; Mismatches 70; Indels 46; Gaps 6;
QY 321 RHFDFDGYNHLVDQNFNPLPDYWLSLYKRLIGPKVLAVHVAGLQKPRGVRIRDKLR 380
Db 12 RQSLIGGNYGLLNTNFTPNPDYYSALIWRQLMGRKALFTTFESGTK-----KIR 60
QY 381 IYAHCTNHHNHNHYVRGSITLFIINLHRSR---KKIKLAGTLRDKLVHQYLLQPY----- 431
Db 61 SYTHCARQSK-----GITVLLMNLNDTTVVAKVELNNSF--SLRHTKHKMSYKRASSQ 112
QY 432 ---GOEG-----LKSksVOLNGOPLVMVDDGTLPELKPPLRAGRTLVIP 473
Db 113 LFGGPNGVIOREEYHLTAKDGNLHSQTMLLNGNALQVNSMGDLPIPIHINSTEPTIA 172
QY 474 PVTMGFFVVKVYNALAC 490
Db 173 PYSIVFVHMRNVVPAC 189
```

Search completed: July 30, 2002, 08:30:45
Job time: 997 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:15:02 ; Search time 53.59 Seconds
(without alignments)
224.703 Million cell updates/sec

Title: US-09-836-461-2_COPY_42_534

Perfect score: 2577
Sequence: 1 GDRPPLPDRAGLKERLLI.....PVTMGFFVKNVNALACRYR 493

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	930.5	36.1	543	2	US-08-922-170B-10
2	930.5	36.1	543	4	US-09-071-739B-2
3	930.5	36.1	543	4	US-09-260-038B-2
4	926.5	36.0	543	4	US-09-181-336-13
5	883.5	34.3	532	4	US-09-181-336-15
6	805	31.2	380	4	US-09-181-336-19
7	779	30.2	380	4	US-09-181-336-17
8	94.5	3.7	685	2	US-08-878-989-1
9	94.5	3.7	685	3	US-09-136-282-2
10	94.5	3.7	685	4	US-09-272-796-1
11	94.5	3.7	685	4	US-09-505-744-2
12	93.5	3.6	934	1	US-08-215-805A-80
13	93	3.6	471	1	US-08-257-341-9
14	91.5	3.6	915	4	US-09-346-237-2
15	91.5	3.6	928	4	US-09-514-539-4
16	91	3.5	1788	2	US-08-962-284-2
17	89.5	3.5	1118	1	US-07-934-393B-2
18	89.5	3.5	1118	1	US-08-278-089A-2
19	89.5	3.5	1118	2	US-08-838-957A-2
20	89	3.5	632	4	US-09-016-080-1
21	89	3.5	1280	2	US-08-583-276-19
22	89	3.5	1280	6	5206352-4
23	88	3.4	1106	1	US-08-435-675B-5
24	88	3.4	1106	1	US-08-336-257A-8
25	87.5	3.4	1122	2	US-08-278-089A-6
26	87.5	3.4	1122	2	US-08-838-957A-6
27	87	3.4	1124	1	US-08-323-474-2

Sequence 2, Appl
Sequence 6, Appl
Sequence 53, Appl
Sequence 53, Appl
Sequence 2, Appl
Patent No. 5386025
Sequence 2, Appl
Sequence 1, Appl
Sequence 7, Appl
Sequence 15, Appl
Sequence 15, Appl
Sequence 2, Appl
Sequence 4, Appl
Patent No. 5268270
Sequence 12, Appl
Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-922-170B-10
; Sequence 10, Application US/08922170B
; Patent No. 5968822
; GENERAL INFORMATION:
; APPLICANT: Iris Pecker, Israel Vlodavsky and Elena
; APPLICANT: Feinstein
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A POLYPEPTIDE
; TITLE OF INVENTION: HAVING HEPARANASE ACTIVITY AND EXPRESSION OF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Robert Sheinbein
; STREET: 2940 Birchtree lane
; CITY: Silver Spring
; STATE: Maryland
; COUNTRY: United States of America
; ZIP: 20906

COMPUTER READABLE FORM:
MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
COMPUTER: Twinhead* Slimnote-890TX
OPERATING SYSTEM: MS DOS version 6.2,
OPERATING SYSTEM: Windows version 3.11
SOFTWARE: Word for Windows version 2.0 converted to
SOFTWARE: an ASCII file
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/922,170B
FILING DATE: 2 SEP 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Friedman, Mark M.
REGISTRATION NUMBER: 33,883
REFERENCE/DOCKET NUMBER: 910/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 972-3-5625553
TELEFAX: 972-3-5625554
TELEX:

INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 543
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-922-170B-10

Query Match 36.1%; Score 930.5; DB 2; Length 543;

```

Best Local Similarity 38.7%, pred. No. 8.5e-90;
Matches 213; Conservative 78; Mismatches 170; Indels 89; Gaps 11;

QY      1  GRRPLPVDRAAGLAEKTLILLDYSTKMPVTNVENFLSLQLQPSLIHD--GWLDFLSSKR 59
Db      22  GLPSGALRPPA--QAQDVLDFTQEPHLVSPSLVTIDANLATDPRFLLIGSPK 79

QY      60  LVTLAGLSPALFRFGKGTDTFLQPNLRNPAKSRGGGPDYILKYNVEDDIVRSVDALDK 119
Db      80  LRTLARGLSPALYLRFGGKTKDFLIP---DPKKESTFEERSYMQSQVNQDI----- 126

QY      120  QKCKIAQ--HPDVMLEQLQREAAQMHVLKLEQF-----SNTYS----- 157
Db      127  ---CKYGSIPDPVEEKLRLWEYPQEQOL--LLREHYOKKFKNSTSRSSVDVLYTFANC SGL 182

QY      158  -----NLIL-----TEPNRYRTVHGRAVNGSOL 180
Db      192  -----NLIL-----TEPNRYRTVHGRAVNGSOL 242

```

Qy	181	GKDYIQKLSLQPIRISYRASISYGNICRPRKNVIALIDGPMKVAGSTVDADVWQHCHYD	240
Db	243	GEDIQIHLKLRK-STFKNAKLYGPDVGQPRKTKAKMLSKFAGGEVIDSVTWHHYLN	301
Qy	241	GRVYKVMDFLTKRLDLTSDQIRKIOKVNVYTPGKKIWLSEGVVTSAGGTNNLSDSYAA	300
Db	302	GRTATREDLPNDVLDIFISSVQKVQFVVESTRPGKVMWLGETSAYGGGAPLLSDTFAA	361
Qy	301	GFLWNTLGLMANOGDIDVIRISFEFDHGYNHVLVDQFNPLPDYWLISLYKRLIGPKVLAV	360
Db	362	GFMWDLKGLUSARNGLEVVRQVFGAGNYHVLVDNFDPDLPDYWLSLLFKKLVGTKVYMA	421
Qy	361	HVAGLQRKPRGPRVIRKURIYACHTNNHHNNHYVRGSTFTLIINLHRSRKKIKUAGTLRD	420
Db	422	SVQGSKRR-----KLRVYLHCTNTONPRYKEGDLTLYAINLHNVTYKRLRPLYPFSN	472
Qy	421	KLVBHQYLQPYGQEGLSKSVOLNGOPLVMWDDGTLPELKPRPLRAGRTLVIPTVTMGFF	480
Db	473	KQVDYILRLRPLGPHGLSKSVQLNGUTLKMWDQTLPLMEKPLRPGSSGLPAFSYSFF	532
Qy	481	VYKNVNALAC	490
Db	533	VIRNAKVAAC	542

RESULT 2
US-09-071-739B-2
; Sequence 2, Application US/09071739B
; Patent No. 6177545
; GENERAL INFORMATION:
; APPLICANT: Iris Pecker et al.
; TITLE OF INVENTION: HEPARANASE SPECIFIC MOLECULAR PROBES
; TITLE OF INVENTION: AND THEIR USE IN RESEARCH AND MEDICAL
; TITLE OF INVENTION: APPLICATIONS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark W. Friedman c/o Anthony Castorina
; STREET: 20001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; OPERATING SYSTEM: Windows version 3.11
; SOFTWARE: word for windows version 2.0 converted to
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,739B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIORITY INFORMATION NUMBER: 09/022,180

Db 362 GFMWLDKLGSLARMGIEVVMRQVFFGAGNHYLVNDFDPLDYWLSLLFKKLVGTVKLMA 421
 QY 361 HVAGLQKRPGRVIRDKLRIYAHCTNNHNNHYVRSITLFIINLHRSRKKIKLAGTLRD 420
 Db 422 SVQGSKR-----KLRVYLHCTNTDNPYKEGDUTLAIINLHNVTKYLRPLPFSN 472
 QY 421 KLVHXYLLQPYGQGLSKSVQNLGQPLVMVDDGTLPKLPRLRAGRTL 470
 Db 473 KQVDKYLRLPLGPHGLLSKSVQNLGLTLKMWDDOTLPPLMEKPLRPGSSL 522

RESULT 6

US-09-181-336-19
 ; Sequence 19, Application US/09181336A
 ; Patent No. 6242238
 ; GENERAL INFORMATION:
 ; APPLICANT: FREEMAN, Craig Geoffrey
 ; APPLICANT: PARISH, Richard
 ; APPLICANT: HAMDORF, Brenton James
 ; APPLICANT: HULETT, Mark Darren
 ; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN
 ; TITLE OF INVENTION: ENDOGLUCURONIDASE AND USES THEREFOR
 ; FILE REFERENCE: 032505-004
 ; CURRENT APPLICATION NUMBER: US/09/181,336A
 ; CURRENT FILING DATE: 1998-10-28
 ; EARLIER APPLICATION NUMBER: AU PP0062
 ; EARLIER FILING DATE: 1997-10-28
 ; EARLIER APPLICATION NUMBER: AU PP0812
 ; EARLIER FILING DATE: 1997-12-09
 ; NUMBER OF SEQ ID NOS: 23
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 19
 ; LENGTH: 380
 ; TYPE: PRT
 ; ORGANISM: Rattus sp.
 US-09-181-336-19

Query Match 31.2%; Score 805; DB 4; Length 380;

Best Local Similarity 47.3%; Pred. No. 9.5e-77;
 Matches 155; Conservative 57; Mismatches 106; Indels 10; Gaps 2;

QY 163 EPNYRTHMGRVAVGSQLGKDYIQLKSLLOPIRIYSRASLYGNIGRPNKVNIALLDGFM 222
 Db 62 EPSEWFKAHISIDGLQGEDFVELHKLQK-SAFQNAKLYGPDIGRGTVKLLRSFL 120
 QY 223 KVAGSTVDVATWQHCYIDGRVVKVMDFLKRLDLSQIRKIQKVVNTYTPGKKIWLEJ 282
 Db 121 KAGGEVDSLTWHYLLNGRVATKEDFLSSDVLDTFILSVQKILKVTREMPGKKWILGE 180
 QY 283 VYTTAGGTNNLSDSYAAAGFLMLNTLGMLANOGIDVYIRHFFDHCYNHLYVDONENPLPD 342
 Db 181 TSSAYGGGAPLSDTFAAGFMWLDKLGSLAQGLIEVVMRQVFFGAGNHYLVNDFDPLD 240
 QY 343 YWLSLLYKRLIGPKVLAVHVGALQKRPGRVIRDKLRIYAHCTNNHNNHYVRSITLFI 402
 Db 241 YWLSLLYKRLIGPKVLAVHVGALQKRPGRVIRDKLRIYAHCTNNHNNHYVRSITLFI 291
 QY 403 INLHRSRKKIKLAGTLRDKLVHXYLLQPYGQGLSKSVQNLGQPLVMVDDGTLPKLPRL 462
 Db 292 LNLHNVTKYLRPLPFSN-----KLRVYLHCTNTDNPYKEGDUTLAIINLHNVTKYLR 351
 QY 463 PLRAGTIVIPVVTGFFVVKVNNALAC 490
 Db 352 PLPAGSSLSVPAFSYGFVIRNAKIAAC 379

RESULT 7

US-09-181-336-17
 ; Sequence 17, Application US/09181336A
 ; Patent No. 6242238
 ; GENERAL INFORMATION:
 ; APPLICANT: FREEMAN, Craig Geoffrey

Db 362 GFMWLDKLGSLARMGIEVVMRQVFFGAGNHYLVNDFDPLDYWLSLLFKKLVGTVKLMA 421
 QY 361 HVAGLQKRPGRVIRDKLRIYAHCTNNHNNHYVRSITLFIINLHRSRKKIKLAGTLRD 420
 Db 422 SVQGSKR-----KLRVYLHCTNTDNPYKEGDUTLAIINLHNVTKYLRPLPFSN 472
 QY 421 KLVHXYLLQPYGQGLSKSVQNLGQPLVMVDDGTLPKLPRLRAGRTL 470
 Db 473 KQVDKYLRLPLGPHGLLSKSVQNLGLTLKMWDDOTLPPLMEKPLRPGSSLGPAFSYFF 532

RESULT 5

US-09-181-336-15
 ; Sequence 15, Application US/09181336A
 ; Patent No. 6242238
 ; GENERAL INFORMATION:
 ; APPLICANT: FREEMAN, Craig Geoffrey
 ; APPLICANT: PARISH, Richard
 ; APPLICANT: HAMDORF, Brenton James
 ; APPLICANT: HULETT, Mark Darren
 ; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN
 ; TITLE OF INVENTION: ENDOGLUCURONIDASE AND USES THEREFOR
 ; FILE REFERENCE: 032505-004
 ; CURRENT APPLICATION NUMBER: US/09/181,336A
 ; CURRENT FILING DATE: 1998-10-28
 ; EARLIER APPLICATION NUMBER: AU PP0062
 ; EARLIER FILING DATE: 1997-10-28
 ; EARLIER APPLICATION NUMBER: AU PP0812
 ; EARLIER FILING DATE: 1997-12-09
 ; NUMBER OF SEQ ID NOS: 23
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 15
 ; LENGTH: 532
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-181-336-15

Query Match 34.3%; Score 883.5; DB 4; Length 532;

Best Local Similarity 38.7%; Pred. No. 7.9e-85;
 Matches 205; Conservative 75; Mismatches 161; Indels 89; Gaps 11;

QY 1 GDRPLPVDRAAGLKEKTLILLDVSINKPNVRYTNENFLSLQDPSIHD-GWLDLSSKR 59
 Db 22 GLSPGALPRPA--QAQDVVDLDFEQEPLHLVSPSFLSVTIDANLATDPRFLLIGSPK 79
 QY 60 LVTLARGLSPAFLEFGKRTDFLQFNLRNPAKSRGCGPDYLYKNVEDDIVRSDVALDK 119
 Db 80 LRTLARGSLPAYLRFEGTKTDFLIF---DPKRETFEERSYQSQVNQDI----- 126
 QY 120 QKGCIAQ-HPDVMLELQREKAAQMLHLVLEKEF-----SNTYS----- 157
 Db 127 --CKYGSIPDVEEKLRLWMPYQEQ-LLREHYQKFKFNKSTYSRSVDVLYTFANCSSL 182
 QY 158 -----NLIL-----TEPNYRTHMGRVAVGSQL 180
 Db 183 DLIFGLNALLRTADLOWNSSNAQLLLDYCSKGYNTISWELGNFSPFKKADIFINGSQL 242
 QY 181 GKDYIQLKSLLOPIRIYSRASLYGNIGRPNKVNIALLDGFMKLVAGSTVDVATWQHCYID 240
 Db 243 GEDFIQLHKLRLK-STFNKAKLYGPDVQGPQRRTAKMLKSLKAGGEVDSVTHWHYLN 301
 QY 241 GRVVKVMDFLKRLDLSQIRKIQKVVNTYTPGKKIWLEGVVTSAGGTNNLSDSYAA 300
 Db 302 GRATREDFLNPVDLIDFISVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 361
 QY 301 GFLWNTLGMLANOGIDVYIRHFFDHCYNHLYVDONENPLPDYWLSLLYKRLIGPKVLAV 360

```
; APPLICANT: PARISH, Richard
; APPLICANT: HAMDRE, Brenton James
; APPLICANT: HULETT, Mark Darren
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN
; TITLE OF INVENTION: ENDOGLUCURONIDASE AND USES THEREFOR
; FILE REFERENCE: 032505-004
; CURRENT APPLICATION NUMBER: US/09/181.336A
; CURRENT FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: AU PP0062
; EARLIER FILING DATE: 1997-10-28
; EARLIER APPLICATION NUMBER: AU PP0812
; EARLIER FILING DATE: 1997-12-09
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Mus musculus
; US-09-181-336-17

Query Match      30.2%; Score 779; DB 4; Length 380;
Best Local Similarity 46.3%; Pred. No. 5.4e-74;
Matches 152; Conservative 56; Mismatches 110; Indels 10; Gaps 2;

Qy 163 EPNYRTHMGRVNGSQGKDYIQLKSLQPIRYSRASLYGNIPGRKPNVIALLDGFM 222
Db 62 EPNSEFWKKAHLIDGLQIGEDFVELHKLQR-SAFQNAKLYGPDIGQPRGKTVKLLRSFL 120
Qy 223 KVAGSTVDATWQHCYIDGRVVKVMDLKLRLDLSQIRKIQKVNTYTPGKKINLEG 282
Db 121 KAGGEVDLSLWHHYHNGRIATDEDFLSSDVLDTFILSVOKILKVTREITPGKKVWLGE 180
Qy 283 VVTSAGGTNNLSYSYAGFLWLNTLGLMANGIDVIRHSHFFDGHYNHLYDONFNPLPD 342
Db 181 TSSAYGGGAPLLSNTFAAGFWMLDKGLSQAQIEVNRQVFFGAGNTHLVNDENFEPLPD 240
Qy 343 YWLSLLYKRLGPKVLAVHVHAGLQKRPGRVIRDKLRIYAHCTNNHNNHYVRSITFLI 402
Db 241 YWLSLLEFKLVGPRVLLSRVAGPD-----RSKLRVYLHCTNVYHPRYQEGDLTLVY 291
Qy 403 INLHRSRKIKLAGTLDKLVHQVLYQPYGOEGLKSKSVOLNGQPLVMVDGTLPELAPR 462
Db 292 LNLHNVTHKLVPPLFRKPPVDTYLLKPSGPDGLLSKSVQLNGQLKMWDEQTLPALTEK 351
Qy 463 PLRAGRTLVIPPTVMGFFVWKNVNALAC 490
Db 352 PLPAGSALSPLAFSYGFFVIRDAKTAAC 379

RESULT 8
US-08-878-989-1
; Sequence 1, Application US/08878989
; Patent No. 5885803
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl G.
; APPLICANT: Lal, Preeti
; APPLICANT: Goli, Surya K.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN
; TITLE OF INVENTION: KINASES
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/878,989
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0321 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 685 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HUVEENOBO1
; CLONE: 39043
; US-08-878-989-1

Query Match      3.7%; Score 94.5; DB 2; Length 685;
Best Local Similarity 23.1%; Pred. No. 0.52;
Matches 84; Conservative 46; Mismatches 144; Indels 89; Gaps 18;

Qy 88 RNPAKSRGGPGPDYLYKKNYEDDIVRSVALDKQCKTAQHPDVMLELQREKAAQ----- 142
Db 84 RKGVLGKG-----PAKCYE-----MTDLTNKVTAAKIPIHSRVAKPHQREKIDKEIELH 134
Qy 143 -----MHLVLLKEQFSNTYSNLIITPEPNYRTMHGRAVNGSOLGKDYIQLKSLQPI-IRI 196
Db 135 RILHKKHVQVHYHFDKENIYILLEYCSRSM-----AHILK-----ARKVLTPEPVY 184
Qy 197 YSRASLYGNIPGRKPNVIALIDGFMKVAGSTVDATWQHCYIDGRVVKVMDFLKTRLLD 256
Db 185 YLRQIVSGLKYLHEQE-----ILHRLDKLGNFIN-----EAMELKVGDFGLAARLE 231
Qy 257 TLDQIRKIQKVNTYTP-----GKIKWLEGVY--TTSAG-----GTNNLSDSY-- 298
Db 232 PLEHRRRTICGTPNYLSPEVLNKQHGCESDIAGLCVMTMLLGRPPFTTNLKETYRC 291
Qy 299 --AAGFLWLNTL-----GMLANQ-----GIDVIRHSHFFDGHY--NHLYDONFNPL 340
Db 292 IREARYTPWSPLLAPAKHLIASMJSKNPDRPSLDDIIRHDDFFLQGFDPDRLSSSCCHTV 351
Qy 341 PDYWLSLLYKRLIGPKVLAVHVHAGLQKRP-----PGRVIRDKLRIYAHCTNNHNNHYVRG 396
Db 352 PDFHLSLSPAKNFF-KKAAALFGGKKDKARYIDITHNRYSKEDEDIY-----KLRLDLAKT 405
Qy 397 SIT 399
Db 406 SIT 408

RESULT 9
US-09-136-282-2
; Sequence 2, Application US/09136282
; Patent No. 6063609
; GENERAL INFORMATION:
; APPLICANT: ANDERSON, KAREN
; APPLICANT: JACKSON, JEFFREY
; APPLICANT: HANSBURY, MICHAEL
; APPLICANT: NERURKAR, SANDHYA
```

Tue Jul 30 10:12:05 2002

```

; APPLICANT: ROSHAK, AMY
; APPLICANT: BOUZYK, MARK
; TITLE OF INVENTION: HUMAN SERUM INDUCIBLE KINASE (SNK)
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,282
; FILING DATE: 20-AUG-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/056,112
; FILING DATE: 20-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Prestia, Paul F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH-70231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0700
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 685 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-09-136-282-2

```

```

Query Match      3.7%; Score 94.5; DB 3; Length 685;
Best Local Similarity 23.1%; Pred. No. 0.52;
Matches 84; Conservative 46; Mismatches 144; Indels 89; Gaps 18;

QY 88 RNPASRGPGPDYLLKNVEDDIVRSVALDKQCKIAQHPDVMLEQREKAAQ-----142
DB 84 RGKVLGKGG-----FAKCYE-----MTDLTNKKVYAAKIIPHSRVAKPHOREKIDKEIELH 134
QY 143 -----MHLVLLKEQFSNTYSNLILTEPNNYRTMHGRAVNGSOLGKDYIQLKSLQP-TRI 196
DB 135 RILHHKHVVQFYHYFEDKENIYILLEYCSRSM-----AHILK---ARKVLTPEVRY 184
QY 197 YSRASLYGNIGRPNRKNTIALDGMFKVAGSTVDATVQHCYIDGRVVKVMDFLKTRLLD 256
DB 185 YLRQIVSGLKYLHEQE-----ILHRDLKLGNNFFIN-----EAMELVGDFGLAARLE 231
QY 257 TILSDQIRKIQKVNTYTP-----GKKIWLGVV--TTSAG---GTNNLSDSY-- 298
DB 232 PLSHRRRTICGTPNLYSPVLNKGQHCSDIWLGCVMYTMLLGPPPTETNLKETYRC 291
QY 299 --AAGFLWLNTL-----GMLANQ-----GIDVVIHRSFFDHGY--NHLVDQNFNPL 340
DB 292 IRARYTMSLLAPAKHLLIASMLSNKPNDRPSLDIIIRHDFLQGTPTDRLSSSCCHTV 351
QY 341 PDWLSLLYKRLGPKVLAVHVGAKRQPR-----PGRVTRDKLRIYVAHTNHHNNHYVG 396
DB 352 PDFHLSSPAKNFT-KKAAALFGGKKDKARYIDTHNRVSKEDEDIY-----KLHDLKKT 405
QY 397 SIT 399
DB 406 SIT 408

```

```

RESULT 10
US-09-272-796-1
; Sequence 1, Application US/09272796
; Patent No. 6207148
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl G.
; APPLICANT: Lal, Preeti
; APPLICANT: Goli, Surya K.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/272,796
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/878,989
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0321 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 685 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HUVENOB01
; CLONE: 39043
;
US-09-272-796-1

```

```

Query Match      3.7%; Score 94.5; DB 4; Length 685;
Best Local Similarity 23.1%; Pred. No. 0.52;
Matches 84; Conservative 46; Mismatches 144; Indels 89; Gaps 18;

QY 88 RNPASRGPGPDYLLKNVEDDIVRSVALDKQCKIAQHPDVMLEQREKAAQ-----142
DB 84 RGKVLGKGG-----FAKCYE-----MTDLTNKKVYAAKIIPHSRVAKPHOREKIDKEIELH 134
QY 143 -----MHLVLLKEQFSNTYSNLILTEPNNYRTMHGRAVNGSOLGKDYIQLKSLQP-TRI 196
DB 135 RILHHKHVVQFYHYFEDKENIYILLEYCSRSM-----AHILK---ARKVLTPEVRY 184
QY 197 YSRASLYGNIGRPNRKNTIALDGMFKVAGSTVDATVQHCYIDGRVVKVMDFLKTRLLD 256
DB 185 YLRQIVSGLKYLHEQE-----ILHRDLKLGNNFFIN-----EAMELVGDFGLAARLE 231
QY 257 TILSDQIRKIQKVNTYTP-----GKKIWLGVV--TTSAG---GTNNLSDSY-- 298

```


Db 232 PLEHRRRTICGTPNYLSPEVLNKGHGCESDIWALGCVMYTMLGRPPFTTNLAKETYRC 291

Qy 299 --AAGFLWLNTL-----GMLANQ-----GIDVIRHSPFDHGY--NHLVDQNFNPL 340

Db 292 IREARYTMPSSLLAPAKHLIASMLSKNEDRPSLDDIIRHDFLQGFPPDRLSSSCCHTV 351

Qy 341 PDYWLSLLYKRLIGPKVLAVHAGLQKRP-----PCRVIRDKLRIYAHCTNHHNNHNYVRG 396

Db 352 PDFHLSSPAKNFF-KKAAALFGGKKKARYIDTHNRVSKEDDIY-----KLRLDLKKT 405

Qy 397 SIT 399

Db 406 SIT 408

RESULT 11

US-09-505-744-2

; Sequence 2, Application US/09505744

; Patent No. 6245544

; GENERAL INFORMATION:

; APPLICANT: Karen M. Anderson

; APPLICANT: Mark M. Bouzyk

; APPLICANT: Michael J. Hansbury

; APPLICANT: Jeffrey R. Jackson

; APPLICANT: Sandhya S. Nerurkar

; APPLICANT: Amy K. Roshak

; TITLE OF INVENTION: HUMAN SERUM INDUCIBLE KINASE (SNK)

; FILE REFERENCE: GH-70231-D1

; CURRENT APPLICATION NUMBER: US/09/505,744

; CURRENT FILING DATE: 2000-02-16

; EARLIER FILING DATE: 1998-08-20

; EARLIER APPLICATION NUMBER: 60/056,112

; EARLIER FILING DATE: 1997-08-20

; NUMBER OF SEQ ID NOS: 3

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 685

; TYPE: PRT

; ORGANISM: HOMO SAPIENS

US-09-505-744-2

Query Match 3.7%; Score 94.5; DB 4; Length 685;

Best Local Similarity 23.1%; Pred. No. 0.52;

Matches 84; Conservative 46; Mismatches 144; Indels 89; Gaps 18;

Qy 88 RNPASRGSGPGDYLLKNYEDDIVSDVALDKQCKIAQHPDVMLELQREKAAQ----- 142

Db 84 RGVLRGKG-----FAKYE-----MTDLTNKKVYAAKIIPHSRVAKPHQREKIDKEIELH 134

Qy 143 -----MHLVLLKEQFSNTYSNLITPEPNRYTMHGRAVNGSOLGKDYIQLKSLQP-TRI 196

Db 135 RLTHKHVVQFHYEDKENIYILLECYSSRM-----AHILK---ARKVLTEPEVRY 184

Qy 197 YSRASLYGPNTRPKKNYIALDGPMPKVGAGTVDAVTWQHCVYIDGRVVKVMDFLKTRLID 256

Db 185 YLRQVSGLKYLHEQE---ILHRDLKLGFFIN-----EAMELKVGDFGLAARLE 231

Qy 257 TSLDQIRKIQKVVNTYTP-----GKKIWLGEV--TTSAG-----GTNNLSDSY-- 298

Db 232 PLEHRRRTICGTPNYLSPEVLNKGHGCESDIWALGCVMYTMLLGRPPFTTNLAKETYRC 291

Qy 299 --AAGFLWLNTL-----GMLANQ-----GIDVIRHSPFDHGY--NHLVDQNFNPL 340

Db 292 IREARYTMPSSLLAPAKHLIASMLSKNEDRPSLDDIIRHDFLQGFPPDRLSSSCCHTV 351

Qy 341 PDYWLSLLYKRLIGPKVLAVHAGLQKRP-----PCRVIRDKLRIYAHCTNHHNNHNYVRG 396

Db 352 PDFHLSSPAKNFF-KKAAALFGGKKKARYIDTHNRVSKEDDIY-----KLRLDLKKT 405

Qy 397 SIT 399

Db 406 SIT 408

Db 406 SIT 408

RESULT 12

US-08-215-805A-80

; Sequence 80, Application US/08215805A

; Patent No. 5559008

; GENERAL INFORMATION:

; APPLICANT: Chang, Yung-Fu

; TITLE OF INVENTION: LEUKOTOXIN GENE FROM PASTEURELLA

; TITLE OF INVENTION: SUIS

; NUMBER OF SEQUENCES: 84

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Nixon, Hargrave, Devans & Doyle

; STREET: Clinton Square, P.O. Box 1051

; CITY: Rochester

; STATE: New York

; COUNTRY: USA

; ZIP: 14603

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC Compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/215.805A

; FILING DATE: 22-MAR-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Timian, Susan J.

; REGISTRATION NUMBER: 34,103

; REFERENCE/DOCKET NUMBER: 19603/61 (D-1329A)

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (716) 263-1636

; TELEFAX: (716) 263-1600

; INFORMATION FOR SEQ ID NO: 80:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 934 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; ORIGINAL SOURCE:

; ORGANISM: Pasteurella suis

; STRAIN: 5943

; IMMEDIATE SOURCE:

; LIBRARY: P. suis DNA in Bacteriophage lambda-dash

; CLONE: (Lambda)yfc33-37

US-08-215-805A-80

Query Match 3.6%; Score 93.5; DB 1; Length 934;

Best Local Similarity 24.3%; Pred. No. 1.1;

Matches 64; Conservative 30; Mismatches 80; Indels 89; Gaps 13;

Qy 101 YLKNYEDDIVSDVALDKQCKIAQHPDVMLELQREK-----AAOMHLVLLKEQFSNT 155

Db 43 YIPKDYEDSGRNGQLDLVKAEE-----DLGIEVQREERNGIATAQNSLSTIQTILGFS 97

Qy 156 YSNLITPEPNRYTMHGRAVNGSOLGKDYIQLKSLQPIRYISRASLYGPNTRPKKNYI 215

Db 98 ERGVVLSAP-----QLDKLLQYKI-SKAPGSSENVAKNLGNQA 135

Qy 216 ALLDGFMPKVGAGTVDAVTWQHCVYIDGRVVK-----VMDFLKTRL-----L 255

Db 136 TLLSGIQLSLGSMAGMD-----LD-ETLKNKGSELDLAKAGLELTNSLIENIANSVQTL 189

Qy 256 DTLSQDI-----RKIQKVVNTYTPGKK-----IWLEGVWTTAGGTNN--LSDSY 298

Db 190 DTFSEQISQGTGKLVQKNGKGLTGLDKLNFSGFGSKAGIGLEVISGLLSGATAALVLADKN 249

Qy 299 A-----AGFLWLNTLGLMANQ 314

Db 250 ASTDRKVGAGF-----ELANQ 265

RESULT 13
US-08-257-341-9
; Sequence 9, Application US/08257341
; Patent No. 5525491
; GENERAL INFORMATION:
; APPLICANT: HUSTON, JAMES S
; APPLICANT: OPPERMANN, HERMANN
; APPLICANT: TIMASHEFF, SERGE N
; TITLE OF INVENTION: SERINE RICH PEPTIDE LINKER
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CREATIVE BIOMOLECULES, INC./PATENT DEPT.
; STREET: 35 SOUTH STREET
; CITY: HOPKINTON
; STATE: MA
; COUNTRY: USA
; ZIP: 01748
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,341
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/842,149
; FILING DATE:
; APPLICATION NUMBER: US 07/662,226
; FILING DATE: 27-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL ESO, PAULA A
; REGISTRATION NUMBER: 32,503
; REFERENCE/DOCKET NUMBER: CRP-064CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000 (ATTY)
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 471 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-257-341-9

Query Match 3.6%; Score 93; DB 1; Length 471;
Best Local Similarity 19.2%; Pred. NO. 0.4;
Matches 90; Conservative 71; Mismatches 152; Indels 156; Gaps 21;

QY 14 LKEKTLILLDVSTKPNVTVNENFLSLQDPSIIHDGWLDFLSSKRLVTLARGLS----- 68
Db 68 LKDRINIVLSRELKEPR--GAHFLAKSLDUALRIEQPELASKYDMYVGGSSVYQEA 125
QY 69 ---PAFLREGGKRTDFLOFQNLNRPKASRGSGPDYILKNYEDDIIVRSDVALDKOGCKI 125
Db 126 MNQPGHLRLFVTR-----INQEFESDFFFEIDLKGYK--LL 160
QY 126 AQHPDVMLEQREKAAQMHVLKLEQFSNTYSNLILTEPNRYTMHGRAVNSQLGKDYI 185
Db 161 PEYPGVLSVEQEGKIKYKFEYVERKKDAS-----VTVSSGSSSSSSSSSGSMIEQD-- 212
QY 186 QLSKLLQPIRIYRSASLSPGPIGRKKNVIAL--DGFPMKVAGSTVDVATWQHCYIDGRV 243
Db 213 -----GLHAGSPAVERLFGVDWAQQTIGCS--DAAVFR--LSAQGRP 252
QY 244 VKVMDFLKRLLDLSQDQIRKIQKVVNTYTFGKKIWLGVVTTTSAGGTNNLSDSYAAGFL 303
Db 253 VL---FVKTDL-----SGALNELQDE--AARLS 275

QY 304 WLNTLMLANQIDVYVIRHSFFDGHYNHLVDQNFENPLPDYWLSLLYKRLIGPKVLAVHVA 363
Db 276 WLATTGVPCAAVLDVVT-----EAGRD-----W--LLGGEVPGDQLSSHLA 315
QY 364 GLQRKPRPGRVTRDKLRIY-----AHCTNHHNNHNVYRGSITLFIINLHRSRKKIKLAGTL 418
Db 316 PAEKV-----SIMADAMRRLLHTLDPATCPFDHQAQH-----RIERARTMEAGLVD 361
QY 419 RDKLVHXYLLQPYGOEGL-----KSKSVQLNGQPLVYV--DDGTLP 459
Db 362 QDDDEEH-----OGLAPAEFLARLAMPDGEDLVVTHGDACLPNI 403

RESULT 14
US-09-346-237-2
; Sequence 2, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:
; APPLICANT: Bisgard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629.200-US
; CURRENT APPLICATION NUMBER: US/09/346,237A
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: PA 1998 00868
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: 60/094,353
; EARLIER FILING DATE: 1998-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Bacillus deramificans
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(915)
; OTHER INFORMATION: Pullulanase
US-09-346-237-2

Query Match 3.6%; Score 91.5; DB 4; Length 915;
Best Local Similarity 23.0%; Pred. NO. 1.7;
Matches 79; Conservative 38; Mismatches 127; Indels 99; Gaps 19;

QY 2 DRRPLPVDRAAGLKE--KTLILLDVSTKNP--VRTVNENFLSLQDPSIIHDGWLDFLSS 57
Db 424 DVRFSDIPNSGMKNKGKYLALTEKGTGPDNVKT-----GID--SL 463
QY 58 KRLVTLARGLSPAFRLFSGKRTDFLOFONLRNPAKSRGGPGDYVYLYKNYEDDIVRSDVAL 117
Db 464 KQLGTHVQLMPVFNASNSVDETD-----PTQDNWGYDP-----RNY--DVPEGQYAT 508
QY 118 DKQCKGKTAQHPDVMLEQREKAAQMHVLKLEQFSNTYSNLILTEPN--NYRTMH--GRAV 175
Db 509 NANGNARIKEFEKEMVLSLREHIGVNMDDVYVNHTEFATQISDFDKIVPEYVYRTDAGNYT 568
QY 176 NGSQGLKDYIQKLSLLQPIRIYRSASLSPGPIGRKKNVIALLDGFMKVAGSTVDVATW 234
Db 569 NGSGTGNEIAERPMVQKF-----IIDSLLKYW 595
QY 235 QHCY-IDGRVYVVMDFLKRLLDLSQDQIRKIQKVVNTYTPGKKIWLGVVTTTSAGGTNN 293
Db 596 VNEYHIDGFRDLMALLGK---DTMS-----KAASELHAINPG--IALLYGEPTW--GGTSA 644
QY 294 LSDSYAAGFLWNLTLGMLANQIDVIRHSFFDGHYNHLVDQN 336
Db 645 LPDDQ-----LLTKG--AQKMGVAV-----FNDNLNRLNLDGN 675

RESULT 15
US-09-514-599-4
; Sequence 4, Application US/09514599

; Patent No. 6350599
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Pullulanase Variants And Methods For
; FILE REFERENCE: 6072.200-US
; CURRENT APPLICATION NUMBER: US/09/514,599
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: PA 2000 00045
; PRIOR FILING DATE: 2000-01-12
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Bacillus deramificans
US-09-514-599-4

Query Match 3.6%; Score 91.5; DB 4; Length 928;
Best Local Similarity 23.0%; Pred. No. 1.8;
Matches 79; Conservative 38; Mismatches 127; Indels 99; Gaps 19;

Qy 2 DRRLPVDRAGLKE--KTLILLDVSTKNP--VRTVNENFLSLQLDPSIIHDGWLDFLSS 57
Db 437 DVRFSDPNSGMKNKGKYLALTEKGTGKPDNVKT-----GID--SL 476

Qy 58 KRLVTLARGLSPAFLRFGGKRTDFLQFONLRNPAKSRGGPGDPYVYKKNVEDDIVRSDVAL 117
Db 477 KQLGITHVQLMPVFNASNSVDETD-----PTQDNNGYDP---RNY--DVPEGQYAT 521

Qy 118 DKQKGCKIAQHPDVWLELQREKAQMHLVLLKEQFSNTYSNLIITEPN-NYRTMH-GRAY 175
Db 522 NANGNARIKEFKEMVLSLHREHIGVNMDDVYVNHFTATQISDFDKIVPEYYRTDDAGNYT 581

Qy 176 NGSQIGKDYIQKSLQPIRIYSRASLYGPNIGRPNKVNIALLDGFMKVGSTVDVAVT-W 234
Db 582 NGSGTGNEIAAERPVMQKF-----IDSLSKYW 608

Qy 235 QHCY-IDGRVYKVMDFLKRLLDLSQIRKIQKVNTYTPGKKIWLKGVVTTTSAGGTNN 293
Db 609 VNEYHIDGFRDLMALLCK---DTMS-----KAASELHAINFG--IALYGEPTW--GGTSA 657

Qy 294 LSDSYAGFLWNLTLGMLANOGIDVYIRHSFFDHGYNHLVDON 336
Db 658 LPDDQ-----LLATKG--AQKGMGVAV-----FNDNLRNALDGN 688

Search completed: July 30, 2002, 08:15:04
Job time: 286 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:14:02 ; Search time 141.47 Seconds
(without alignments)
317.982 Million cell updates/sec

Title: US-09-836-461-2_COPY_130_534
Perfect score: 2129
Sequence: 1 NPAKSRGGPGDYLLKNYED.....PVTMGFFVKNVNALACRYR 405

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*

1: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Description	
Result No.	Score	Match	Length DB ID		
1	2129	100.0	534 22 AAB85216	Heparanase-like pr	
2	2122	99.7	492 22 AAB84664	Amino acid sequenc	
3	2090	98.2	592 22 AAU07424	Human heparanase-1	
4	2090	98.2	592 22 AAB85215	Heparanase-like pr	
5	2090	98.2	592 22 AAY97632	Human heparanase,	
6	2080	97.7	592 22 AAB81062	Human Heparanase-2	
7	1829	85.9	538 22 AAY97633	Human Heparanase,	
8	1822	85.6	439 22 AAU07423	Human heparanase-1	
9	1822	85.6	480 22 AAU07418	Novel human extrac	
10	1822	85.6	480 22 AAB85217	Heparanase-like pr	
11	1822	85.6	480 22 AAY97634	Human heparanase,	

12	1112	52.2	214	22	AAM99905	Human excretory re
13	1112	52.2	214	22	AAAM3704	Human bladder anti
14	805	37.8	380	20	AAAY17085	Rat heparanase enz
15	799.5	37.6	543	20	AAAY02345	A human heparanase
16	799.5	37.6	543	21	AAAB08849	Amino acid sequenc
17	799.5	37.6	543	21	AAAY57590	Human heparanase.
18	799.5	37.6	543	21	AAAY52990	Human heparanase p
19	799.5	37.6	543	22	AAAY97635	Human heparanase p
20	799.5	37.6	592	20	AAAY02346	A human heparanase
21	799.5	37.6	592	21	AAAB08850	Amino acid sequenc
22	796.5	37.4	530	20	AAAY34173	Human pre-prohepar
23	796.5	37.4	543	22	AAAB88361	Human membrane or
24	795.5	37.4	543	20	AAAY17082	Human heparanase e
25	795.5	37.4	543	20	AAAB86206	Human heparanase i
26	795.5	37.4	588	20	AAAY30124	A human protein wi
27	790	37.1	535	21	AAAB08851	A murine heparanas
28	779	36.6	380	20	AAAY17084	Mouse heparanase e
29	778	36.5	156	22	AAG65963	Human heparanase-1
30	752.5	35.3	532	20	AAAY17083	Seq ID No: 15 of W
31	658	30.9	262	22	AAAM24147	Human EST encoded
32	410	19.3	488	22	AAAB31469	Amino acid sequenc
33	388	18.2	488	22	AAAB31470	Amino acid sequenc
34	383	18.0	488	22	AAAB31472	Amino acid sequenc
35	375	17.6	488	22	AAAB31471	Amino acid sequenc
36	278	13.1	137	22	AAG65961	Human heparanase-1
37	278	13.1	159	22	AAG65964	Human heparanase-1
38	270	12.7	112	22	AAU07425	Human heparanase-1
39	247	11.6	256	21	AAG13479	Arabidopsis thalia
40	203	9.5	118	22	AAG65962	Mouse heparanase-1
41	162	7.6	115	22	AAAB85218	Mouse heparanase-1
42	155	7.3	935	22	ABB69219	Drosophila melanog
43	144	6.8	38	20	AAAY34186	Human pre-prohepar
44	127	6.0	24	22	AAB85220	Heparanase-like pr
45	127	6.0	25	22	AAB85221	Heparanase-like pr

ALIGNMENTS

RESULT		1
AAB85216		
ID	AAB85216 standard; Protein; 534 AA.	
XX	AC	
XX	AAB85216;	
XX	07-SEP-2001 (first entry)	
DE	Heparanase-like protein Hpa2 splice variant #2.	
KW	Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;	
KW	cytostatic; neuroprotective; cerebroprotective; immunosuppressive;	
KW	antipsoriatic; nootropic; antiinflammatory; antiarthritic; antiasthmatic;	
KW	antidiabetic; antiarteriosclerotic; vulnery.	
OS	Homo sapiens.	
XX	XX	
XX	WO200146392-A2.	
PD	28-JUN-2001.	
XX	XX	
PF	21-DEC-2000; 2000WO-GB04963.	
PR	22-DEC-1999; 99GB-0030392.	
XX	07-APR-2000; 2000GB-0008713.	
PA	(OXFO-) OXFORD GLYCOSCIENCES UK LTD.	
XX	XX	
PI	McKenzie EA, Stamps AC, Terrett JA, Tyson KL;	
XX	WPI; 2001-418056/44.	
DR	N-PSDB; AAH22672.	
XX	XX	
PT	Novel homologs of heparanase, present in three splice variants, useful	

PT	for identifying agents that modulate heparanase, useful in the treatment and/or prophylaxis of abnormal levels of heparanase	
XX		
PS	Claim 1; Fig 2; 97pp; English.	
XX	The invention provides a homologue to heparanase which is present in three splice variants. The heparanase homologue polypeptide is useful in the treatment of a human or non-human animal or for use in diagnosis. Vectors comprising the heparanase homologue polynucleotides are useful in the transformation or transfection of a prokaryotic or eukaryotic host. The modulators of the polypeptide are useful in the manufacture of a medicament for the treatment and/or prophylaxis of a condition/disease associated with abnormal levels of the heparanase homologue, including cancer, central nervous system (CNS) and neurodegenerative diseases, cardiovascular diseases such as restenosis following angioplasty and atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus, allografts, inflammatory diseases, arthritis, vascular restenosis, tumour growth and progression, asthma, Alzheimer's disease, diabetic retinopathy, wound healing and inflammation. The polypeptide is also useful in diagnosis and research. The present sequence represents the amino acid sequence of the mid-sized splice variant of the heparanase-like protein Hpa2 of the invention.	
XX	Sequence 534 AA;	
SQ		
PT	Query Match	100.0%; Score 2129; DB 22; Length 534;
XX	Best Local Similarity	100.0%; Pred. No. 5.5e-220;
PS	Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
XX		
XX	1 NPAKSRGGPGDYLLKNEYEDIVRSVDALDKOKGCKIAQHDPVMLELOREKAAQHMLVLL 60	
CC		
CC	130 npaks:ggpgpdyllkneyedivrsdvaldkqgckiaqhpdmvmlgrekaqmhlvll 189	
CC		
QY	61 KEQFSNTYSNLIITPEPNYRTMHGRAVNSQLGKDYIQLKSLLOPIRIYSRASLYGPNIG 120	
CC		
CC	190 keqfsntysnliitepnnyrtmhgravngsqlgkdyiqklsllqprlyrsrasygnlig 249	
CC		
QY	121 RPRKNVIALDGMKVGSTVDVATWQHGYIDGRVVKVMDFLKTRLLDLSQIRKIQRV 180	
CC		
CC	250 rprknvialldgmkvagstvdatwqhcyidgrvvkvmfdiktrlldtldsqirkiqkv 309	
CC		
QY	181 VNTYTPGCKIWLKGVVTTTSAGTNNLSDSYAAGFLMLNTLGLMANGIDVIRHSEFFDHG 240	
CC		
CC	310 vntytpgkklwlegvvttsaggttnnlsdsyaagflwntlglmlangidvvrhseffdhg 369	
CC		
QY	241 YNHLVDQNFPLPDYWLISLLYKRLIGPKVLAVHVGALQRPGRVIRDKLRIYAHCTNH 300	
CC		
CC	370 ynhlvdqnfplpdywlsillykrligpkvlavhvaglqrprgprvirdklriyahctnh 429	
CC		
QY	301 HHNNYVRGSSITLFTINLHRSRKKIKLAGTLRDKLVHGYLLQPYGQEGLSKSVQLNGQPL 360	
CC		
CC	430 hnnhnyvrsgstlftinlhrsrrkkiklagtlrdklvhoyllqpygqeglsksvqlngqpl 489	
CC		
QY	361 VMVDDGTLPELKPRLRAGRTLPIPPVTMGFFVVKVNNALACRYR 405	
CC		
CC	490 vmvddgtlpeclkprlragrtlpippvmtgmfvvknnvalacryr 534	
CC		
RESULT 2		
AAB84664		
ID	AAB84664 standard; Protein; 492 AA.	
XX		
AC	AAB84664;	
XX		
DT	05-SEP-2001 (first entry)	
XX		
DE	Amino acid sequence of human heparanase-like polypeptide.	
XX	Human; heparanase-like polypeptide; gene therapy; cancer; angiogenesis;	
KW	trauma; autoimmune disease; skin disease; cardiovascular disease;	
KW	nervous system disease; inflammation; arthritis; genitalia;	
KW	male fertility; erectile dysfunction.	

Db 388 hnmvyrsgitlfiinlhrxrkklagtlrdklvqhyllyqpygqeglsksvqlngqpl 447

Qy 361 VMVDDGTLPELKPRLRAGRTLVIPPTVMGFVVKVNNALACRYR 405
|||||

Db 448 vmvddgtlpeikprlragrtlvipptvmgfyvkvnnalacryr 492

RESULT 3

ID AAU07424

XX AC AAU07424;

XX DT 18-DEC-2001 (first entry)

XX DE Human heparanase-like protein splice variant #1.

XX KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic; antiproliferative; cardiac; vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide; fungicide; ophthaimological; extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis; hyperproliferative disorder; neoplasm; cardiovascular disorder; cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection; nervous system disorder; Alzheimer's disease; ocular disorder; sunburn; wound healing; food additive; heparanase.

XX OS Homo sapiens.

XX PN WO200179253-A1.

XX PD 25-OCT-2001.

XX PF 11-APR-2001; 2001WO-US11643.

XX PR 18-APR-2000; 2000US-198123P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX FI Flscella M, Shi Y, Ebner R, Ruben SM;

XX DR WPI; 2001-611720/70.

XX DR N-PSDB; AAS13848.

XX PT New nucleic acids encoding extracellular matrix polypeptides, for diagnosing, treating, preventing or ameliorating human disorders and disease, such as, autoimmune, hyperproliferative or cardiovascular disorders -

XX PS Disclosure; Page 14; 308pp; English.

CC The invention relates to novel isolated polynucleotides (I) encoding extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by (I) are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. The antibodies to the polypeptides can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis,

CC hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. The present sequence represents the amino acid sequence of human heparanase-like protein, splice variant #1.

XX

SQ Sequence 592 AA;

Query Match 98.2%; Score 2090; DB 22; Length 592;
Best Local Similarity 87.5%; Pred. No. 1e-215; 0; Indels 58; Gaps 1;
Matches 405; Conservative 0; Mismatches 0;

QY 1 NPAKSRGPGPDYILKNYEDDIVRSDVALDKQKGAIAQHPDVMLEQREKAQMHVL 60
|||||

Db 130 npaksergpgpdyilknyeddivrsdvaldkqkgckiaqhpdmlelqrekaqmhlvl 189
|||||

QY 61 KEQFSNTYSNLILT----- 74
|||||

Db 190 keqfsntysnllltarsldklynfadcsghlflalnlnrrnnpnswssalsllkysa 249
|||||

QY 75 -----EPNNYRTMHGRAVNGSOLGKDYIQLKSLLOPIRIYSRASLYGPNIGRP 122
|||||

Db 250 skkyniswelgnepnnyrtmhgravnsglqkdyiqllsqipriyrsaslygpnigrp 309
|||||

QY 123 RKNVIALLDGFMKVGASTVDVATWQHICYIDGRVVKVMDFLKTRLDLSQIRKIQKVVN 182
|||||

Db 310 rknavialldgfmkvagstvdavtwhqcyidgrvkvvmdfiktrlldtldsqirkiqkvvn 369
|||||

QY 183 TYPGKKIWLGVVTTSGAGTNNLSDSYAAGFLWLNTLGLANOGIDVIVIRHSFFDHGYN 242
|||||

Db 370 tytpgkkiwlegvvttsgagtnnlsdsyaagflwntlglmlanqgidvivrhsffdhgyn 429
|||||

QY 243 HLVDQNFNLPDYWLISLLYKRLIGPKVLAVHVAGLQKRPGRVIRDKLRIYAHCTNNHN 302
|||||

Db 430 hlvdqnfnpdywlislllykrligpkvlavhvagiqrkprgprvirdklriyahcnnhnn 489
|||||

QY 303 HNYVRGSITLFIINLHRSRKKIKLAGTLRDKLVHQLLOPYGQGLKSKSVQLNGQPLVM 362
|||||

Db 490 hnyvrgsitlfiinlhrrs-kkiklagtlrdklvqhyllyqpygqeglsksvqlngqplvm 549
|||||

QY 363 VDDGTLPELKPRLRAGRTLVIPPTVMGFVVKVNNALACRYR 405
|||||

Db 550 vddgtlpeikprlragrtlvipptvmgfyvkvnnalacryr 592
|||||

RESULT 4

AAB85215

ID AAB85215 standard; Protein; 592 AA.

XX AC AAB85215;

XX DT 07-SEP-2001 (first entry)

XX DE Heparanase-like protein Hpa2 splice variant #1.

XX KW Heparanase; splice variant; homologue; heparanase-like protein; Hpa2; cytostatic; neuroprotective; cerebroprotective; immunosuppressive; antipsoriatic; neutropic; antiinflammatory; antiarthritic; antiasthmatic; antidiabetic; antiarteriosclerotic; vulnerary.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT Misc-difference 237 /label= unknown

XX FT /note= "encoded by ANC"

XX PN WO200146392-A2.

XX PD 28-JUN-2001.

XX PF 21-DEC-2000; 2000WO-GB04963.

XX PR 22-DEC-1999; 99GB-0030392.

XX PR 07-APR-2000; 2000GB-0008713.

XX PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.

XX

PI McKenzie EA, Stamps AC, Terrett JA, Tyson KL;
XX WPI; 2001-418056/44.
DR N-PSDB; AAH22671.
XX
PT Novel homologs of heparanase, present in three splice variants, useful
PT for identifying agents that modulate heparanase, useful in the
PT treatment and/or prophylaxis of abnormal levels of heparanase
XX
PS
PS Claim 1; Fig 1; 97pp; English.
XX
CC The invention provides a homologue to heparanase which is present in
CC three splice variants. The heparanase homologue polypeptide is useful in
CC the treatment of a human or non-human animal or for use in diagnosis.
CC Vectors comprising the heparanase homologue polynucleotides are useful in
CC the transformation or transfection of a prokaryotic or eukaryotic host.
CC The modulators of the polypeptide are useful in the manufacture of a
CC medicament for the treatment and/or prophylaxis of a condition/disease
CC associated with abnormal levels of the heparanase homologue, including
CC cancer, central nervous system (CNS) and neurodegenerative diseases,
CC cardiovascular diseases such as restenosis following angioplasty and
CC atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus,
CC allografts, inflammatory diseases, arthritis, vascular restenosis,
CC tumour growth and progression, asthma, Alzheimer's disease, diabetic
CC retinopathy, wound healing and inflammation. The polypeptide is also
CC useful in diagnosis and research. The present sequence represents the
CC amino acid sequence of the largest splice variant of the heparanase-like
XX protein Hpa2 of the invention.
XX
SQ Sequence 592 AA;

Query Match 98.2%; Score 2090; DB 22; Length 592;
Best Local Similarity 87.5%; Pred. No. 1e-215;
Matches 405; Conservative 0; Mismatches 0; Indels 58; Gaps 1;

QY 1 NPAKSRGGPGPDYLLKNYEDDIVRSVDALDKQKCKIAQHPDVMLELQREKAAQHLVLL 60
DB 130 npakrrggpgpdyllknyeddivrsdvaldkqgckiaqhpdvmlqlrekaqmhlvl 189
QY 61 KEQFSNTYSNLIIT----- 74
DB 190 keqfsntysnliitarsldklynfadcsglhlfalnrrpnpswxssalsllkysa 249
QY 75 -----EPNNYRTMHGRAVNGSOLGKDYIQLKSLLOPIRIYSRASLYGPNIGRP 122
DB 250 skkynlswelgnepnyrtmhgravnsglqkdyiqklsllqpiriyraslygpnigrp 309
QY 123 RKNVIALLDGFMKVGAGSTVDVAVTWHQCYIDGRVVKVMDFLKTRLDLTLSQIRKIQKVVN 182
DB 310 rknvialldgfmkvagstvavtwhqcyidgrvvkvmdflktrlldtldsqirkikqvv 369
QY 183 TYTPGKKIWLGVVTTTSAGTNNLSDSYAGFLWLNTLGLMANOGIDVVIRHSFFDHGYN 242
DB 370 tytpgkklwlegvvtttsagtnnlsdsyaagflwntlgmlanogidvvirhsffdhgyn 429
QY 243 HLVDQNFNPLPDYWLSSLYKRLIRLPGKVLAVHVAGLQKRPGRVIRDKLRIYAHCTNNHN 302
DB 430 hlvdqnfnpdpdywlsllkrlirpgrpvlavhvaglrqrpgrvirdklriyahctnnhn 489
QY 303 HNVYRGSITLFIINLHRSRKKIACGLPLRDLKLVHQYLLQYGGEGLSKSVQINGPLVM 362
DB 490 hnyvrgsitlfiinlhrrsrkkiaclglplrdklvhyqllqyggeglsksvqingqplvm 549
QY 363 VDDGTLPELPRPLRACRTLVIPVTMGFFVKNVNALACRYR 405
DB 550 vddgtlpeelprrlragrtlvipvtmgffvknvnalacryr 592

RESULT 5
ID AAY97632
ID AAY97632 standard; Protein; 592 AA.
XX

AC AAY97632;
XX
DT 20-APR-2001 (first entry)
XX
XX Human heparanase, hnhpl, protein sequence.
DE
XX Heparanase: hnhpl; wound healing; angiogenesis; restenosis; Scrape;
KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
KW gene therapy; human.
XX
XX Homo sapiens.
OS
XX WO200100643-A2.
PN
XX 04-JAN-2001.
PD
XX 19-JUN-2000; 2000WO-IL00358.
PF
XX 25-JUN-1999; 99US-0140801.
PR
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PA Pecker I, Michal I, Itzhaki H;
PI
XX WPI; 2001-137930/14.
DR N-PSDB; AAA91097.
DR
XX New polynucleotides and polypeptides that are distantly homologous to
PT heparanase, useful in wound healing, as well as in gene therapy
PT protocols for angiogenesis, restenosis, atherosclerosis, or
PT inflammation
PT
XX Claim 10; Fig 1; 67pp; English.
PS
XX This sequence represents a heparanase of the invention.
CC The heparanase DNA and protein sequences are useful in wound healing,
CC angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary
CC diseases, neurodegenerative diseases (such as Scrape, Alzheimer's
CC disease, and Creutzfeldt-Jakob disease) or viral infections. The
CC heparanase coding sequence is particularly useful in gene therapy.
XX
XX Sequence 592 AA;
SQ

Query Match 98.2%; Score 2090; DB 22; Length 592;
Best Local Similarity 87.5%; Pred. No. 1e-215;
Matches 405; Conservative 0; Mismatches 0; Indels 58; Gaps 1;

QY 1 NPAKSRGGPGPDYLLKNYEDDIVRSVDALDKQKCKIAQHPDVMLELQREKAAQHLVLL 60
DB 130 npakrrggpgpdyllknyeddivrsdvaldkqgckiaqhpdvmlqlrekaqmhlvl 189
QY 61 KEQFSNTYSNLIIT----- 74
DB 190 keqfsntysnliitarsldklynfadcsglhlfalnrrpnpswnssalsllkysa 249
QY 75 -----EPNNYRTMHGRAVNGSOLGKDYIQLKSLLOPIRIYSRASLYGPNIGRP 122
DB 250 skkynlswelgnepnyrtmhgravnsglqkdyiqklsllqpiriyraslygpnigrp 309
QY 123 RKNVIALLDGFMKVGAGSTVDVAVTWHQCYIDGRVVKVMDFLKTRLDLTLSQIRKIQKVVN 182
DB 310 rknvialldgfmkvagstvavtwhqcyidgrvvkvmdflktrlldtldsqirkikqvv 369
QY 183 TYTPGKKIWLGVVTTTSAGTNNLSDSYAGFLWLNTLGLMANOGIDVVIRHSFFDHGYN 242
DB 370 tytpgkklwlegvvtttsagtnnlsdsyaagflwntlgmlanogidvvirhsffdhgyn 429
QY 243 HLVDQNFNPLPDYWLSSLYKRLIRLPGKVLAVHVAGLQKRPGRVIRDKLRIYAHCTNNHN 302
DB 430 hlvdqnfnpdpdywlsllkrlirpgrpvlavhvaglrqrpgrvirdklriyahctnnhn 489
QY 430 hlvdqnfnpdpdywlsllkrlirpgrpvlavhvaglrqrpgrvirdklriyahctnnhn 489
DB

Qy 303 HNYVRSITLFIINLHRSRKKIKIAGTLRDKLVHQLQYQGGELKSKSVOLNGQPLVM 362
|||||
Db 490 hnyvrgsitlfiinlhrrskkiklagtlrklvghyllqpygqeglkksvqlngqplvm 549
|||||
Qy 363 VDDGTLPELKPRLRAGRTLVIPTVMGFFVVKVNNALACRYR 405
|||||
Db 550 vddgtlpeikprlragrtlviptvmgffvkvnnalacryr 592
|||||

RESULT 6
AAB81062
ID AAB81062 standard; Protein; 592 AA.
XX AAB81062;
XX
XX
DT 20-JUN-2001 (first entry)
XX
XX Human Heparanase-2 amino acid sequence.
XX
KW Heparanase 2; human; endoglucuronidase; heparan sulphate; metastasis;
neocangiogenesis; vaccine; autoimmune disorder; blood coagulation;
KW cancer; diabetes; ischaemia; sepsis; stroke; cardiovascular; thrombosis.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 156..169
FT /label= Immunogenic_epitope
FT Region 249..262
FT /label= Immunogenic_epitope
FT Region 505..518
FT /label= Immunogenic_epitope
XX
XX WO200121814-A1.
XX
XX 29-MAR-2001.
XX
XX 11-SEP-2000; 2000WO-EP08837.
XX
XX 23-SEP-1999; 99EP-0118805.
PR 07-JUL-2000; 2000EP-0114649.
XX
XX (MERE) MERCK PATENT GMBH.
XX
XX Duecker K, Sirrenberg C;
XX
XX WPI: 2001-308089/32.
DR N-PSDB; AAF86101.
XX
XX New heparanase-2 polypeptide useful in diagnosing (the susceptibility
of a subject to) and as vaccines against e.g. autoimmune disorders, or
PT cardiovascular disease, cancer, diabetes, ischemia, sepsis, stroke, or
PT thrombosis
XX
XX Claim 1; Page 42-43; 46pp; English.
XX
XX This invention relates to a human heparanase-2 protein and the cDNA
sequence encoding it. Heparanase-2 is a member of the endoglucuronidase
CC family of polypeptides and it degrades heparan sulphate proteoglycans and
CC HSPGs (ubiquitous macromolecules of cell surfaces, basement membranes and
CC the extracellular matrix). HSPGs support the vascular endothelium and
CC stabilise the structure of the capillary wall. Heparanases may be
CC associated with neocangiogenesis and metastasis related to malignant
CC tumour formation. Heparanase-2 polynucleotides and proteins are useful as
CC vaccines for inducing an immunological response against autoimmune
CC disorders, blood coagulation disorders, cancer, diabetes, ischaemia,
CC sepsis, stroke, cardiovascular diseases, or thrombosis, as well as in
CC diagnosing (the susceptibility of a subject to) these diseases.
CC Heparanase-2 fragments may be used as immunogens to produce antibodies
CC immunospecific to the polypeptides, and to identify membrane bound
CC soluble receptors, agonists or antagonists that compete with the binding
CC of the polypeptide to the receptors. An antibody specific for
CC heparanase-2 can be used in the diagnosis of the above diseases and in

CC isolating or identifying clones expressing heparanase-2. The present
CC sequence represents heparanase-2. Three regions of heparanase-2 with high
CC immunogenicity (immunogenic epitopes) can be used to raise antibodies
CC against heparanase-2.
XX
SQ Sequence 592 AA;
Query Match 97.7%; Score 2080; DB 22; Length 592;
Best Local Similarity 87.0%; Pred. No. 1.2e-214;
Matches 403; Conservative 1; Mismatches 1; Indels 58; Gaps 1;
Qy 1 NPAKSRGGPGDYLLKNEYEDDIVRSDVALDKOKGCKIAQHPDVMLELQREKAAQHLLVLL 60
|||||
Db 130 npakserggpgdyllkneyeddivrsdvaldkqkgckiaqhpdmvlvlgrekaaqmhlvl 189
|||||
Qy 61 KEQFSNTYSNLILT----- 74
|||||
Db 190 keqfsntysnlltarsldklynfadcsglhlifalnrlrnpnswssalsilkysa 249
|||||
Qy 75 -----EPNNYRTMHGRAVNGSOLGKDYIQLKSLLOPIRIYSRASLYGPNIGRP 122
|||||
Db 250 skkyniswelgnepnnyrtmhgravngsqkgkdyqlksllqpiriysraslygpnigrp 309
|||||
Qy 123 RKNVIALLDGFMKVAGSTVDVAVTWQHCYIDGRVVKVMDFLKTRLLDTLSDQIRKIQKVVN 182
|||||
Db 310 rknvialldgfmkvagstvdavtwqhcyidgrvvkvmdfklrlltdltsdqirkikgvn 369
|||||
Qy 183 TYTPGKKIWLGVVTTTSAGTNNLSDSYAAGFLWLNTGLMLANOGIDVVIRISFFDHGVN 242
|||||
Db 370 tytpgkkilwlgvvttsaggtnnlsdsyaagflwntgmlanogidvvirhsffdhgyn 429
|||||
Qy 243 HLVDQNFNPLPDYWLSSLVKRLIGPKVLAVHVAGLQKRPGRVIRDKLRIYAHCTNHHN 302
|||||
Db 430 hlvdqnfnpdpdywlsllvkrligpkvlavhvaglqkprgrvirdklriyahctnhhn 489
|||||
Qy 303 HNYVRSITLFIINLHRSRKKIKLAGTLRDKLVHQLQYQGGELKSKSVOLNGQPLVM 362
|||||
Db 490 hnyvrgsitlfiinlhrrskkiklagtlrklvghyllqpygqeglkksvqlngqplvm 549
|||||
Qy 363 VDDGTLPELKPRLRAGRTLVIPTVMGFFVVKVNNALACRYR 405
|||||
Db 550 vddgtlpeikprlragrtlviptvmgffvkvnnalacryr 592
|||||

RESULT 7
AAY97633
ID AAY97633 standard; Protein; 538 AA.
XX
XX AAY97633;
XX
XX 20-APR-2001 (first entry)
XX
XX Human heparanase, hnhpl pn9 form, protein sequence.
XX
XX Heparanase; hnhpl; wound healing; angiogenesis; restenosis; Scrape;
KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
XX gene therapy; human.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT Misc-difference 305 /note= "encoded by GAC"
XX
XX WO200100643-A2.
XX
XX 04-JAN-2001.
XX
XX 19-JUN-2000; 2000WO-IL00358.
XX
XX 25-JUN-1999; 99US-0140801.
PR

XX	PA	(INSI-) INSIGHT STRATEGY & MARKETING LTD.	
XX	PI	Pecker I, Michal I, Itzhaki H;	
XX	XX	WPI; 2001-137930/14.	
DR	DR	N-PSDB; AAA91098.	
XX	XX	New polynucleotides and polypeptides that are distantly homologous to	
PT	PT	heparanase, useful in wound healing, as well as in gene therapy	
PT	PT	protocols for angiogenesis, restenosis, atherosclerosis, or	
PT	PT	inflammation -	
XX	XX	Claim 10; Page 61-62; 67pp; English.	
PS	PS	This sequence represents a heparanase of the invention.	
XX	CC	The heparanase DNA and protein sequences are useful in wound healing,	
CC	CC	angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary	
CC	CC	diseases, neurodegenerative diseases (such as Scrape, Alzheimer's	
CC	CC	disease, and Creutzfeldt-Jakob disease) or viral infections. The	
CC	CC	heparanase coding sequence is particularly useful in gene therapy.	
XX	XX	Sequence 538 AA;	
SQ	SQ		
Query Match 85.9%; Score 1829; DB 22; Length 538;			
Best Local Similarity 88.3%; Pred. No. 1.2e-187;			
Matches 364; Conservative 2; Mismatches. 36; Indels 10; Gaps 4;			
QY	1	NPAKSRGGPGDPYLLKNYEDDIVRSDVALDKQKCKIAQHPDVMLEQREKA-----AQM 55	
Db	130	npaksrpgpdyllknyed--arsldklynfadcs-gihllifalnairnpnswns 186	
QY	56	HLVLLKEQESNTY--SNLIITPEPNYRTMHGRAVNSQLGKDYIQLKSLLOPIRYSRAS 113	
Db	187	alsllkysaskynisweignepnnyrtmhgravnsgqgkdyiqklsllqpirlystras 246	
QY	114	LYGPNIGRPRKPNVIALLDGFMKVGASTVDVATWQHICYIDGRVVKVMDFLKRLDLSQ 173	
Db	247	lygpnigrprkpnvialldgfmkvagstvdtwqhcyidgrvkvmdfiktllldtlsa 306	
QY	174	IRKIQKVVNTYTPGKKIWLVEGVVTTGAGTNNLSDSYAGFLWNTLGLMANQIDVVR 233	
Db	307	irkikvntytpgkklwlegvvttsaggtnnlsdsyagflwntlglmanqgidvvr 366	
QY	234	HSFDFHGYNHLVDQNFPLDYLLKRLIGPKVLAVHAGLQKRPGRVIRDKLRI 293	
Db	367	hsfddghnhlvdqnfpldyllskrligpkvlavhagqlrkprgvrirdklri 426	
QY	294	YAHCTNHHNHYVRGSTITLFIINLHRSRKKIKLAGTLRDKLVHQYLLQPYQEGLSKSV 353	
Db	427	yahctnhhnhvrgstltfiinlhrrskkiklagtlrdklvqhyllyqpygqeglsksv 486	
QY	354	QLNQCPQLVWDDGTLPELKPRLRAGTLYIPVTWGEFFVVKVNALACRYR 405	
Db	487	qlngqpvlwddgtlpeikprlragnetlylvpvtwgmffvkvvnalacryr 538	
RESULT 8			
AAU07423	ID	AAU07423 standard; Protein; 439 AA.	
XX	AC	AAU07423;	
XX	XX	18-DEC-2001 (first entry)	
XX	DE	Human heparanase-like protein.	
XX	XX	Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;	
KW	KW	antiproliferative; cardiant; vasotropic; cerebroprotective; neotropic;	
KW	KW	neuroprotective; antibacterial; virucide; fungicide; ophthalmological;	
KW	KW	extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;	
KW	KW	hyperproliferative disorder; neoplasm; cardiovascular disorder;	
KW	KW	cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;	
KW	KW	nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;	
XX	XX	wound healing; food additive; heparanase.	
OS	OS	Homo sapiens.	
XX	XX	WO200179253-A1.	
PN	PN	25-OCT-2001.	
XX	XX	11-APR-2001; 2001WO-US11643.	
PF	PF	18-APR-2000; 2000US-198123P.	
XX	XX	(HUMA-) HUMAN GENOME SCI INC.	
XX	PA	Fiscella M, Shi Y, Ebner R, Ruben SM;	
PI	PI	WPI; 2001-611720/70.	
XX	XX	New nucleic acids encoding extracellular matrix polypeptides, for	
XX	XX	diagnosing, treating, preventing or ameliorating human disorders and	
PT	PT	disease, such as, autoimmune, hyperproliferative or cardiovascular	
PT	PT	disorders -	
PT	PT	Dislosure; Page 13-14; 308pp; English.	
XX	XX	The invention relates to novel isolated polynucleotides (I) encoding	
XX	XX	extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by	
CC	CC	(I) are used to prevent, treat or ameliorate a medical condition in e.g.	
CC	CC	humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They	
CC	CC	are also used in diagnosing a pathological condition or susceptibility to	
CC	CC	a pathological condition. The antibodies to the polypeptides can also be	
CC	CC	used in alleviating symptoms associated with the disorders and in	
CC	CC	diagnostic immunoassays e.g. radioimmunoassays or enzyme linked	
CC	CC	immunosorbant assays (ELISA). Disorders which are diagnosed or treated	
CC	CC	include autoimmune diseases e.g. rheumatoid arthritis,	
CC	CC	hyperproliferative disorders e.g. neoplasms of the breast or liver,	
CC	CC	cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders	
CC	CC	e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.	
CC	CC	Alzheimer's disease, infections caused by bacteria, viruses and fungi	
CC	CC	and ocular disorders e.g. corneal infection. The polypeptides can also	
CC	CC	be used to aid wound healing and epithelial cell proliferation, to	
CC	CC	prevent skin aging due to sunburn, to maintain organs before	
CC	CC	transplantation, for supporting cell culture of primary tissues, to	
CC	CC	regenerate tissues and in chemotaxis. The polypeptides can also be used	
CC	CC	as a food additive or preservative to increase or decrease storage	
CC	CC	capabilities. The present sequence represents the amino acid sequence	
CC	CC	of human heparanase-like protein.	
XX	XX	Sequence 439 AA;	
SQ	SQ		
Query Match 85.6%; Score 1822; DB 22; Length 439;			
Best Local Similarity 86.7%; Pred. No. 4.7e-187;			
Matches 351; Conservative 0; Mismatches 0; Indels 54; Gaps 1;			
QY	1	NPAKSRGGPGDPYLLKNYEDDIVRSDVALDKQKCKIAQHPDVMLEQREKAQMHLVLL 60	
Db	89	npaksrpgpdyllknyed----- 108	
QY	61	KEQFSNTYSNLIITPEPNYRTMHGRAVNSQLGKDYIQLKSLLOPIRYSRASLYGPNIG 120	
Db	109	-----epnyrtmhgravnsglqgkdyiqklsllqpirlyrsaslygpnig 154	
QY	121	RPRKNVIALLDGFMKVGASTVDVATWQHICYIDGRVVKVMDFLKRLDLSQIRKIQV 180	
Db	155	rprknvialldgfmkvagstvdtwqhcyidgrvvkvmdfiktllctisqirkiqkv 214	
QY	181	VNTYTPGKKIWLVEGVVTTGAGTNNLSDSYAGFLWNTLGLMANQIDVIRHSEFFDHG 240	
Db	215	vntytpgkklwlegvvtttsaggtnnlsdsyagflwntlglmanqidvvrhseffdhg 274	

QY 241 YNHLVDQNFNPLDYWLSLYKRLIGPKVLAVHVAGLQRPQRPVIRDKLRIYAHCTNH 300
Db 275 ynhlvdqnfnpdywlslykrligpkvlavhvaglrkprprvirdklriyahctnh 334
QY 301 HNNHYVSGSTWTFIINLHRSRKKIKLAGTLRDKLVHVLQPYGOEGLKSKSVOLNGOPL 360
Db 335 hnnhyvrgstlfiinlhrsrrkkiklagtlrdklvvhvlylqpygqglksksvqlngqpl 394
QY 361 VMVDDGTLPELKPRLPRAGRTLVIPTVTMGFFVVKNNVALACRYR 405
Db 395 vmvddgtlpeikprlpragrtlviptvtmgffvkvknvalacryr 439
RESULT 9
AAU07418
ID AAU07418 standard; Protein; 480 AA.
XX
AC AAU07418;
DT 18-DEC-2001 (first entry)
XX
DE Novel human extracellular matrix (ECM) protein #1.
XX
KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW antiproliferative; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; cerebral ischemia; infection;
KW nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;
KW wound healing; food additive.
XX
OS Homo sapiens.
XX
XX WO200179253-A1.
PN
XX
XX 25-OCT-2001.
PD
XX
XX 11-APR-2001; 2001WO-US11643.
XX
XX 18-APR-2000; 2000US-198123p.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Fiscella M, Shi Y, Ebner R, Ruben SM;
XX
XX WPI; 2001-611720/70.
DR
XX N-PSDB; AAS13843.
XX
XX New nucleic acids encoding extracellular matrix polypeptides, for
PT diagnosing, treating, preventing or ameliorating human disorders and
PT disease, such as, autoimmune, hyperproliferative or cardiovascular
PT disorders -
XX
XX Claim 1; Page 292-293; 308pp; English.
PS
XX
XX The invention relates to novel isolated polynucleotides (I) encoding
CC extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by
CC (I) are used to prevent, treat or ameliorate a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility to
CC a pathological condition. The antibodies to the polypeptides can also be
CC used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before

CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities. The present sequence represents the amino acid sequence
CC of novel human extracellular matrix (ECM) protein #1.
XX
SQ Sequence 480 AA;
Query Match 85.6%; Score 1822; DB 22; Length 480;
Best Local Similarity 86.7%; Pred. No. 5.5e-187;
Matches 351; Conservative 0; Mismatches 0; Indels 54; Gaps 1;
QY 1 NPAKSRGGPGPDYLLKNYEDDIVRSVALDKQKCKIAQHPDVMLELOREKAAQMHVLVL 60
Db 130 npaksggpgpdyllknyed----- 149
QY 61 KEQFSNTYSNLLITPNNYRTMHGRAVNGSQLGKDYIQLKSLLOPIRISRASLYGPNIG 120
Db 150 -----epnnnyrtmhgravngsqigkdyiglkslilqpirisrasygpnig 195
QY 121 RPRKNVIALLDGFMKVGAGSTVDVATWQHCYIDGRVVKVMDFLKTLRLDTLSQIRKIOKV 180
Db 196 rprknvialldgfmkvagstvdavtcqhcyidgrvkvkvmdfllktrlldtlsdqirkikqv 255
QY 181 VNTYTPGKKIWLEGVYVTTTSAGGTNNLSDSYAAGFLWLNTLGLANQGDVIVIRHSFFDHG 240
Db 256 vntytpgkkiwlegvvttsaggtnnlsdsyaagflwntlglmangqdvivirhsffdhg 315
QY 241 YNHLVDQNFNPLDYWLSLYKRLIGPKVLAVHVAGLQRPQRPVIRDKLRIYAHCTNH 300
Db 316 ynhlvdqnfnpdywlslykrligpkvlavhvaglrkprprvirdklriyahctnh 375
QY 301 HNNHYVSGSTWTFIINLHRSRKKIKLAGTLRDKLVHVLQPYGOEGLKSKSVOLNGOPL 360
Db 376 hnnhyvrgstlfiinlhrsrrkkiklagtlrdklvvhvlylqpygqglksksvqlngqpl 435
QY 361 VMVDDGTLPELKPRLPRAGRTLVIPTVTMGFFVVKNNVALACRYR 405
Db 436 vmvddgtlpeikprlpragrtlviptvtmgffvkvknvalacryr 480
RESULT 10
AAB85217
ID AAB85217 standard; Protein; 480 AA.
XX
AC AAB85217;
XX
DT 07-SEP-2001 (first entry)
XX
DE Heparanase-like protein Hpa2 splice variant #3.
XX
KW Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;
KW cytostatic; neuroprotective; cerebroprotective; immunosuppressive;
KW antipsoriatic; nootropic; antiinflammatory; antiarthritic; antiasthmatic;
KW antidiabetic; antiarteriosclerotic; vulnery.
XX
OS Homo sapiens.
XX
XX WO200146392-A2.
PN
XX
PD 28-JUN-2001.
XX
PF 21-DEC-2000; 2000WO-GB04963.
XX
PR 22-DEC-1999; 99GB-0030392.
PR 07-APR-2000; 2000GB-0008713.
XX
PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI McKenzie EA, Stamps AC, Terrett JA, Tyson KL;
XX WPI; 2001-418056/44.
DR

DR N-PSDB; RAH22673.

XX Novel homologs of heparanase, present in three splice variants, useful

PT for identifying agents that modulate heparanase, useful in the

PT treatment and/or prophylaxis of abnormal levels of heparanase -

XX

XX

PS Claim 1; Fig 3; 97pp; English.

XX

CC The invention provides a homologue to heparanase which is present in

CC three splice variants. The heparanase homologue polypeptide is useful in

CC the treatment of a human or non-human animal or for use in diagnosis.

CC Vectors comprising the heparanase homologue polynucleotides are useful in

CC the transformation or transfection of a prokaryotic or eukaryotic host.

CC The modulators of the polypeptide are useful in the manufacture of a

CC medicament for the treatment and/or prophylaxis of a condition/disease

CC associated with abnormal levels of the heparanase homologue, including

CC cancer, central nervous system (CNS) and neurodegenerative diseases,

CC cardiovascular diseases such as restenosis following angioplasty and

CC atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus,

CC allografts, inflammatory diseases, arthritis, vascular restenosis,

CC tumour growth and progression, asthma, Alzheimer's disease, diabetic

CC retinopathy, wound healing and inflammation. The polypeptide is also

CC useful in diagnosis and research. The present sequence represents the

CC amino acid sequence of the smallest splice variant of the heparanase-

CC like protein Hpa2 of the invention.

XX

XX Sequence 480 AA;

Query Match' 85.6%; Score 1822; DB 22; Length 480;

Best Local Similarity 86.7%; Pred. No. 5.5e-187;

Matches 351; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

QY 1 NPAKSRGGGPDYLLKNYEDDIDVRSVDALDKQKGCKIAQHDPVMLELQREKAAQMHVLVL 60

DB 130 npaksggpgpdyllknedy----- 149

QY 61 KEQFSNTYSNLITLTPNNVMTMHGRAVNGSQLGKDYIOLKSLLOPIRIYSRASLYGPNIG 120

DB 150 -----epnnrytmhgravngsqlgkdyiqklsllqpiriyraslygpnig 195

QY 121 RPRKNVIALLDGFMKVAGSTVDVAVTWQHCYIDGRVVKVMDFLKRLDPTLSQIRKIKV 180

DB 196 rprknvialldgfmkvagstvdavtwqhcylgrvvkvmvdfklrldptlsdqirkikqv 255

QY 181 VNTYTPGKKIWLLEGVVTTSAGCTNNLSDSYAAGFLWNTLGLMANGQIDVIRHSHFFDHG 240

DB 256 vntytpgkklwlegvvvttsagctnnlsdsyaagflwntlglmanqgidvirhshffdhg 315

QY 241 YNHLVDQNFNPLPDYWSLLYKRLIGPKVLAVHVAGLQRPGRVIRDKLRIYAHCTNH 300

DB 316 ynhlvdqnfnpldpywsllykrlgpkvlavhvagqlrkprgvrirdklriyahctnh 375

QY 301 HHNNYVRSITLFIINLHRSRKKIKLAGTLRDKLVHQYLLQYPYQEGLSKSVQLNGQPL 360

DB 376 hbnnyvrgsitlfiinlhrsrrkkiklagtlrdklvhqyllqpygqegllksksvqlngqpl 435

QY 361 VMVDDGTLPELKPRLRAGRTLVIPTVTMGFFVVKVNNALACRYR 405

DB 436 vmvddgtlpeikprlragrtlivipvtmgffvkvnnalacryr 480

RESULT 11

AA97634

ID AAY97634 standard; Protein; 480 AA.

XX

AC AAY97634;

XX

XX 20-APR-2001 (first entry)

XX

DE Human heparanase, hnhp1 pn5 form, protein sequence.

XX

XX Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;

atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;

neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;

gene therapy; human.

OS Homo sapiens.

XX

PN WO200100643-A2.

XX

PD 04-JAN-2001.

XX

PF 19-JUN-2000; 2000WO-IL00358.

XX

PR 25-JUN-1999; 99US-0140801.

XX

PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX

XX Pecker I, Michel I, Itzhaki H;

PI

XX WPI; 2001-137930/14.

DR

DR N-PSDB; AAA91099.

XX

XX New polynucleotides and polypeptides that are distantly homologous to

PT heparanase, useful in wound healing, as well as in gene therapy

PT protocols for angiogenesis, restenosis, atherosclerosis, or

PT inflammation -

XX

PS Claim 10; Page 63; 67pp; English.

XX

CC This sequence represents a heparanase of the invention.

CC The heparanase DNA and protein sequences are useful in wound healing,

CC angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary

CC diseases, neurodegenerative diseases (such as Scrape, Alzheimer's

CC disease, and Creutzfeldt-Jakob disease) or viral infections. The

CC heparanase coding sequence is particularly useful in gene therapy.

XX

XX Sequence 480 AA;

Query Match 85.6%; Score 1822; DB 22; Length 480;

Best Local Similarity 86.7%; Pred. No. 5.5e-187;

Matches 351; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

QY 1 NPAKSRGGGPDYLLKNYEDDIDVRSVDALDKQKGCKIAQHDPVMLELQREKAAQMHVLVL 60

DB 130 npaksggpgpdyllknedy----- 149

QY 61 KEQFSNTYSNLITLTPNNVMTMHGRAVNGSQLGKDYIOLKSLLOPIRIYSRASLYGPNIG 120

DB 150 -----epnnrytmhgravngsqlgkdyiqklsllqpiriyraslygpnig 195

QY 121 RPRKNVIALLDGFMKVAGSTVDVAVTWQHCYIDGRVVKVMDFLKRLDPTLSQIRKIKV 180

DB 196 rprknvialldgfmkvagstvdavtwqhcylgrvvkvmvdfklrldptlsdqirkikqv 255

QY 181 VNTYTPGKKIWLLEGVVTTSAGCTNNLSDSYAAGFLWNTLGLMANGQIDVIRHSHFFDHG 240

DB 256 vntytpgkklwlegvvvttsagctnnlsdsyaagflwntlglmanqgidvirhshffdhg 315

QY 241 YNHLVDQNFNPLPDYWSLLYKRLIGPKVLAVHVAGLQRPGRVIRDKLRIYAHCTNH 300

DB 316 ynhlvdqnfnpldpywsllykrlgpkvlavhvagqlrkprgvrirdklriyahctnh 375

QY 301 HHNNYVRSITLFIINLHRSRKKIKLAGTLRDKLVHQYLLQYPYQEGLSKSVQLNGQPL 360

DB 376 hbnnyvrgsitlfiinlhrsrrkkiklagtlrdklvhqyllqpygqegllksksvqlngqpl 435

QY 361 VMVDDGTLPELKPRLRAGRTLVIPTVTMGFFVVKVNNALACRYR 405

DB 436 vmvddgtlpeikprlragrtlivipvtmgffvkvnnalacryr 480

RESULT 12

AA99905

ID	AM99905 standard; Protein; 214 AA.	PR	08-SEP-2000; 2000US-0231414.
XX	AC	PR	08-SEP-2000; 2000US-0232080.
XX	AM99905;	PR	08-SEP-2000; 2000US-0232081.
XX	DT	PR	12-SEP-2000; 2000US-0231968.
XX	07-JAN-2002 (first entry)	PR	14-SEP-2000; 2000US-0232397.
XX	Human excretory related polypeptide SEQ ID NO 642.	PR	14-SEP-2000; 2000US-0232398.
XX		PR	14-SEP-2000; 2000US-0232399.
XX		PR	14-SEP-2000; 2000US-0232400.
KW	Human; nontropic; neuroprotective; cytostatic; dermatological; virucide;	PR	14-SEP-2000; 2000US-0232401.
KW	immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;	PR	14-SEP-2000; 2000US-0233063.
KW	antiparkinsonian; antischlicking; antianaemic; antiarthritic; cancer;	PR	14-SEP-2000; 2000US-0233064.
KW	antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;	PR	14-SEP-2000; 2000US-0233065.
KW	antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;	PR	21-SEP-2000; 2000US-0234223.
KW	antiparasitic; cardiac; immune disorder; cardiovascular disorder;	PR	21-SEP-2000; 2000US-0234274.
KW	neurological disease; infection; nephrotropic; gene therapy; vaccine;	PR	25-SEP-2000; 2000US-0234997.
KW	excretory system.	PR	25-SEP-2000; 2000US-0234998.
XX		PR	26-SEP-2000; 2000US-0235484.
OS	Homo sapiens.	PR	27-SEP-2000; 2000US-0235834.
XX		PR	27-SEP-2000; 2000US-0235836.
PN	WO200155313-A2.	PR	29-SEP-2000; 2000US-0236327.
XX		PR	29-SEP-2000; 2000US-0236367.
PD		PR	29-SEP-2000; 2000US-0236368.
XX		PR	29-SEP-2000; 2000US-0236369.
XX		PR	29-SEP-2000; 2000US-0236370.
PF		PR	02-OCT-2000; 2000US-0236802.
XX	17-JAN-2001; 2001WO-US01323.	PR	02-OCT-2000; 2000US-0237037.
PR	31-JAN-2000; 2000US-0179065.	PR	02-OCT-2000; 2000US-0237038.
PR	04-FEB-2000; 2000US-0180628.	PR	02-OCT-2000; 2000US-0237039.
PR	24-FEB-2000; 2000US-0184664.	PR	02-OCT-2000; 2000US-0237040.
PR	02-MAR-2000; 2000US-0186350.	PR	13-OCT-2000; 2000US-0239935.
PR	16-MAR-2000; 2000US-0189874.	PR	13-OCT-2000; 2000US-0239937.
PR	17-MAR-2000; 2000US-0190076.	PR	20-OCT-2000; 2000US-0240960.
PR	18-APR-2000; 2000US-0198123.	PR	20-OCT-2000; 2000US-0241221.
PR	19-MAY-2000; 2000US-0205515.	PR	20-OCT-2000; 2000US-0241785.
PR	07-JUN-2000; 2000US-0209467.	PR	20-OCT-2000; 2000US-0241786.
PR	28-JUN-2000; 2000US-0214886.	PR	20-OCT-2000; 2000US-0241787.
PR	30-JUN-2000; 2000US-0215135.	PR	20-OCT-2000; 2000US-0241808.
PR	07-JUL-2000; 2000US-0216647.	PR	20-OCT-2000; 2000US-0241809.
PR	07-JUL-2000; 2000US-0216880.	PR	20-OCT-2000; 2000US-0241826.
PR	11-JUL-2000; 2000US-0217487.	PR	01-NOV-2000; 2000US-0244617.
PR	11-JUL-2000; 2000US-0217496.	PR	08-NOV-2000; 2000US-0246474.
PR	14-JUL-2000; 2000US-0218290.	PR	08-NOV-2000; 2000US-0246475.
PR	26-JUL-2000; 2000US-0220963.	PR	08-NOV-2000; 2000US-0246476.
PR	26-JUL-2000; 2000US-0220964.	PR	08-NOV-2000; 2000US-0246477.
PR	14-AUG-2000; 2000US-0224518.	PR	08-NOV-2000; 2000US-0246478.
PR	14-AUG-2000; 2000US-0224519.	PR	08-NOV-2000; 2000US-0246523.
PR	14-AUG-2000; 2000US-0225213.	PR	08-NOV-2000; 2000US-0246524.
PR	14-AUG-2000; 2000US-0225214.	PR	08-NOV-2000; 2000US-0246525.
PR	14-AUG-2000; 2000US-0225266.	PR	08-NOV-2000; 2000US-0246526.
PR	14-AUG-2000; 2000US-0225267.	PR	08-NOV-2000; 2000US-0246527.
PR	14-AUG-2000; 2000US-0225268.	PR	08-NOV-2000; 2000US-0246528.
PR	14-AUG-2000; 2000US-0225270.	PR	08-NOV-2000; 2000US-0246532.
PR	14-AUG-2000; 2000US-0225447.	PR	08-NOV-2000; 2000US-0246609.
PR	14-AUG-2000; 2000US-0225757.	PR	08-NOV-2000; 2000US-0246610.
PR	14-AUG-2000; 2000US-0225758.	PR	08-NOV-2000; 2000US-0246611.
PR	18-AUG-2000; 2000US-0225759.	PR	08-NOV-2000; 2000US-0246613.
PR	22-AUG-2000; 2000US-0226279.	PR	17-NOV-2000; 2000US-0249207.
PR	22-AUG-2000; 2000US-0226681.	PR	17-NOV-2000; 2000US-0249208.
PR	22-AUG-2000; 2000US-0226686.	PR	17-NOV-2000; 2000US-0249209.
PR	22-AUG-2000; 2000US-0227182.	PR	17-NOV-2000; 2000US-0249210.
PR	23-AUG-2000; 2000US-0227009.	PR	17-NOV-2000; 2000US-0249211.
PR	30-AUG-2000; 2000US-0228924.	PR	17-NOV-2000; 2000US-0249212.
PR	01-SEP-2000; 2000US-0229287.	PR	17-NOV-2000; 2000US-0249213.
PR	01-SEP-2000; 2000US-0229343.	PR	17-NOV-2000; 2000US-0249214.
PR	01-SEP-2000; 2000US-0229344.	PR	17-NOV-2000; 2000US-0249215.
PR	01-SEP-2000; 2000US-0229345.	PR	17-NOV-2000; 2000US-0249216.
PR	05-SEP-2000; 2000US-0229509.	PR	17-NOV-2000; 2000US-0249217.
PR	05-SEP-2000; 2000US-0229513.	PR	17-NOV-2000; 2000US-0249218.
PR	06-SEP-2000; 2000US-0230437.	PR	17-NOV-2000; 2000US-0249244.
PR	06-SEP-2000; 2000US-0230438.	PR	17-NOV-2000; 2000US-0249245.
PR	08-SEP-2000; 2000US-0231242.	PR	17-NOV-2000; 2000US-0249264.
PR	08-SEP-2000; 2000US-0231243.	PR	17-NOV-2000; 2000US-0249265.
PR	08-SEP-2000; 2000US-0231244.	PR	17-NOV-2000; 2000US-0249297.
PR	08-SEP-2000; 2000US-0231413.		

PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-024617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.

PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX WPI; 2001-514652/56.
XX N-PSDB; AAI64065.
DR
XX
XX
PT Forty five bladder related polynucleotides, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases -
XX
XX
PS Claim 11; SEQ ID NO 98; 482pp + sequence listing; English.
XX
XX The invention relates to forty five novel bladder related
CC polynucleotides. The polynucleotides and the polypeptides that they
CC encode are useful in the diagnosis, treatment and prevention of:
CC cancer, particularly breast and ovarian cancer, and other cancers
CC of the adrenal gland, bone, bone marrow, breast, gastrointestinal
CC tract, liver, lung, or urogenital system; immune disorders such as
CC Addison's disease, allergies, autoimmune haemolytic anaemia,
CC autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple
CC sclerosis, rheumatoid arthritis and ulcerative colitis; cardiovascular
CC disorders such as myocardial ischaemias; wound healing; neurological
CC diseases such as cerebral anoxia and epilepsy; and infectious diseases
CC such as viral, bacterial, fungal and parasitic infections. Numerous
CC examples of each type of disorder are given in the specification.
CC The polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. The polynucleotides are
CC useful for chromosome identification. They are also useful as probes
CC for diagnosing or treating a disorder related to the female reproductive
CC system, particularly breast and/or ovary cancer. The present
CC sequence is a novel bladder antigen provided in the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 214 AA;

Query Match 52.2%; Score 1112; DB 22; Length 214;
Best Local Similarity 100.0%; Pred. No. 4.5e-111;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 196 VTTAGGTNNLSDSYAAGFLNLTGLMLANQSIDVIRHSFFDHGYNHLVDONFPLDPY 255
Db 5 vttsaggtnnlsdsyaagflnltglmlanqgidvirhsffdhgynhlvdqfnplpdy 64
Qy 256 WLSLLYKRLIGPKVLAVHVAGLQRPGRVIRDKLRIVAHCTNHHNHYVSGSTLFTI 315
Db 65 wlsillykrligpkvlavhvaglrqrpgrvirdkrlriyahctnhhnmhyvrgstlfti 124
Qy 316 NLHRSRKIKLAGTLRDLKLVHQLLPYQEGKLSKSVQLNSQPLVMVDDGTLPELKRPR 375
Db 125 nlhrsrkiklagtlrdklvlhqllypqyqeglkksvqlngqplvmvddgtlpeikprp 184
Qy 376 LRAGRTLVIPTVMGFFVKKVNNALACRYR 405
|||||

Db	185	lragrtlvippvtmgffvkvknvalacryr	214
Db	181	tsaygggapltsdtfaagfmwldkqlsglgltevmrmrgvffgagnyhlvdenfeplpd	240
Qy	255	YWLSLLVKRLIGPKVLAVHVGALQKRPGRVIRDKLRIYAHCTNHHNHYVRGSTITLFI	314
Db	241	ywlsllfkklvgpkvlnsrvgpd-----rsklrlylhtcnvybpcyregdltlyv	291
Qy	315	INLHRSRKKIKLAGTLDRDKLVHQYLLQPYGOEGGLKSKSVQLNGOPLVMVDDGTLPELKPR	374
Db	292	lnlhvntkhlklpppmfmrpvdkyllkpfsgdglksksvqngqtlkmvdeqtlpaitek	351
Qy	375	PLRAGRTLIVPPVTMGFFVKNVALAC	402
Db	352	plpagsslsvpafsygffvknakiac	379
RESULT	15		
AY02345			
ID	AY02345	standard; Protein; 543 AA.	
XX	AC	AY02345;	
DT	09-JUL-1999	(first entry)	
XX	DE	A human heparanase protein.	
XX	KW	Heparanase; hp; modulator; heparin-binding growth factor;	
XX	KW	cellular response; cytokine; cell interaction; plasma lipoprotein;	
XX	KW	cellular susceptibility; infection; disintegration;	
XX	KW	neurodegenerative plaque; wound healing; angiogenesis; restenosis;	
XX	KW	atherosclerosis; inflammation; neurodegenerative disease; neurallise;	
XX	KW	plasma heparin; micrometastasis; autoimmune lesion; renal failure.	
XX	OS	Homo sapiens.	
XX	XX	WO9911798-A1.	
XX	PN	11-MAR-1999.	
XX	PF	31-AUG-1998; 98WO-US17954.	
XX	PR	02-JUL-1998; 98US-0109386.	
XX	PR	02-SEP-1997; 97US-0922170.	
XX	PA	(FRIE/) FRIEDMAN M M.	
XX	PA	(HADA-) HADASIT MEDICAL RES SERVICES & DEV.	
XX	PA	(INSI-) INSIGHT STRATEGY & MARKETING LTD.	
XX	PI	Feinstein E, Pecker I, Vlodaysky I;	
XX	DR	WPI; 1999-302255/25.	
XX	DR	N-PSDB; AAX35648.	
XX	PT	New human polynucleotide useful for treating angiogenesis,	
XX	PT	restenosis, and inflammation	
XX	PS	Claim 6; Fig 1; 63pp; English.	
XX	CC	The specification describes a polypeptide having heparanase (hp)	
XX	CC	activity. The recombinant protein is used as a modulator of	
XX	CC	heparin-binding growth factors, cellular responses to heparin-binding	
XX	CC	growth factors and cytokines, cell interaction with plasma lipoproteins,	
XX	CC	cellular susceptibility to viral, protozoal and bacterial infections	
XX	CC	or disintegration of neurodegenerative plaques. Heparanase may be	
XX	CC	useful for conditions such as wound healing, angiogenesis, restenosis,	
XX	CC	atherosclerosis, inflammation, neurodegenerative diseases, and viral	
XX	CC	infections. Mammalian heparanase can be used to neutralize plasma	
XX	CC	heparin, and anti-heparanase antibodies may be applied for	
XX	CC	immunodetection and diagnosis of micrometastases, autoimmune lesions,	
XX	CC	and renal failure in biopsy specimens, plasma samples, and body fluids.	
XX	CC	The present sequence represents human heparanase.	
XX	Sequence	543 AA;	

Search completed: July 30, 2002, 08:14:03
Job time: 340 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:16:29 ; Search time 71.68 Seconds
(without alignments)
542.915 Million cell updates/sec

Title: US-09-836-461-2_COPY_130_534
Perfect score: 2129
Sequence: 1 NPAKSRGGPGPDYLYKNYED.....PVTMGFFVKNVNALACRYR 405
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1822	85.6	480	2 JC7506	heparanase protein
2	301	14.1	521	2 T45608	hypothetical prote
3	155	7.3	190	2 T01953	hypothetical prote
4	102.5	4.8	1260	2 T14022	reverse transcript
5	101	4.7	535	2 S58740	cytochrome-c oxida
6	100.5	4.7	630	2 C71374	probable glucose i
7	98.5	4.6	335	2 E84992	tryptophan--trna l
8	97	4.6	1020	2 T18342	glutamate dehydrog
9	96	4.5	898	2 B84471	cytoplasmic aconit
10	95	4.5	456	2 T24442	hypothetical prote
11	95	4.5	512	2 T1261	cytochrome-c oxida
12	95	4.5	557	2 D97210	extracellular neut
13	95	4.5	740	2 T51619	probable ethylene
14	94.5	4.4	460	2 G90554	p46-like (mycoplas
15	94	4.4	625	2 G86855	glucose inhibited
16	94	4.4	1517	2 B81393	DNA-directed RNA p
17	93	4.4	320	2 D90478	conserved hypothet
18	92.5	4.3	328	2 AC1210	drtp-D-glucose 4,6
19	92.5	4.3	876	2 G89952	DNA polymerase I
20	92.5	4.3	1616	2 T16600	vitellogenin vit-1
21	92.5	4.3	2241	2 S09811	hypothetical prote
22	92	4.3	455	2 S67627	probable membrane
23	92	4.3	461	1 S65187	GPI-anchor biosynt
24	92	4.3	534	2 S17993	cytochrome-c oxida
25	92	4.3	898	2 T10101	aconitate hydratase
26	91.5	4.3	986	2 F82229	sarcosine oxidase
27	91.5	4.3	986	2 AH3056	sarcosine oxidase
28	91	4.3	1272	2 C90593	hypothetical prote
29	90.5	4.3	283	2 H81435	probable lipoprote

30	90.5	4.3	682	2 A44493	serum-inducible ki
31	90	4.2	650	2 F81654	regulatory protein
32	90	4.2	709	2 T16584	hypothetical prote
33	89	4.2	687	2 A50345	polyphosphate kina
34	89	4.2	770	2 S77523	hypothetical prote
35	89	4.2	2380	2 E71604	hypothetical prote
36	88.5	4.2	286	2 A86576	chromosome partiti
37	88.5	4.2	286	2 B72047	probable fatty-acy
38	88.5	4.2	341	2 D70761	ATP-dependent heli
39	88.5	4.2	933	2 H90247	PRP8 protein homol
40	88.5	4.2	2403	2 T30875	cytochrome c oxida
41	88	4.1	555	2 A52147	hypothetical prote
42	88	4.1	1032	2 S53571	calcium channel pr
43	88	4.1	1106	1 CHRBA2	collagen alpha 3(V
44	88	4.1	3137	1 A37797	hypothetical prote
45	87.5	4.1	411	2 D87377	hypothetical prote

ALIGNMENTS

RESULT 1
JC7506
heparanase protein 2a - human
C:Species: Homo sapiens (man)
C:Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 01-Dec-2000
C:Accession: JC7506
R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hirccock, M.;
Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000
A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase famli
A:Reference number: JC7506
A:Accession: JC7506
A:Molecule type: mRNA
A:Residues: 1-480 <MCK>
A:Cross-references: GB:AF282885
C:Comment: This protein, a intracellular membrane-bound enzyme, has biological and th
therapies.
C:Genetics:
A:Gene: hpa2a
A:Map position: 10q23-10q24
C:Keywords: heparin binding; membrane bound

Query Match 85.6%; Score 1822; DB 2; Length 480;

Best Local Similarity 86.7%; Pred. No. 1.6e-141;

Matches 351; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

QY 1 NPAKSRGGPGPDYLYKNYEDDIVRSDVALDKKQCKIAQHDPVMLELQREKAAQMHVLVL 60

DB 130 NPAKSRGGPGPDYLYKNYED----- 149

QY 61 KEQFSNTYSNLILTEPNNYRTMHGRAVNSQLGKDYIQLKSLLOPTIRIYSRASLYGPNIG 120

DB 150 -----EPNNTRTMHGRAVNSQLGKDYIQLKSLLOPTIRIYSRASLYGPNIG 195

QY 121 RPRKNVIALLDGFMKVAGSTVDATVWQHCYIDGRVVKVMDFLKTRLLDLSQIRKIQKV 180

DB 196 RPRKNVIALLDGFMKVAGSTVDATVWQHCYIDGRVVKVMDFLKTRLLDLSQIRKIQKV 255

QY 181 VNTYTPGKKIWLEGVVTTSAGGTNNLSDSYAAAGFLWNLGLMLANOGIDVTRHSPFDHG 240

DB 256 VNTYTPGKKIWLEGVVTTSAGGTNNLSDSYAAAGFLWNLGLMLANOGIDVTRHSPFDHG 315

QY 241 YNHLVDQNFPLPDYVLSLLYKRLIGPKVLAVHVGAKRQPRGVRIRDKLRIYAHCTNH 300

DB 316 YNHLVDQNFPLPDYVLSLLYKRLIGPKVLAVHVGAKRQPRGVRIRDKLRIYAHCTNH 375

QY 301 HHNHTVRSGITLFIINLHRSRKKIKLAGTLRDKLVHQYLLQYPYQEGELKSKSVQLNGQPL 360

DB 376 HHNHTVRSGITLFIINLHRSRKKIKLAGTLRDKLVHQYLLQYPYQEGELKSKSVQLNGQPL 435

QY 361 VMVDDGTLPDLKPRPLRAGRTLVPVPTWFFVVKVKNVNALACRYR 405

|||||

Db 436 VWDDGTLPELKPRLRAGRTLVIPVMTGFFVVKVNNALACRYR 480

RESULT 2

T45608

hypothetical protein F13G24.30 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000

C:Accession: T45608

R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Voickaert, G.; Ba

submitted to the Protein Sequence Database, December 1999

A:Reference number: 223009

A:Accession: T45608

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-521 <BEV>

A:Cross-references: EMBL:AL133421

A:Experimental source: cultivar Columbia; BAC clone F13G24

C:Genetics:

A:Map position: 5

A:Introns: 53/3; 66/1; 127/2; 177/1; 256/1; 319/2; 361/2; 394/3

A:Note: F13G24.30

Query Match 14.1%; Score 301; DB 2; Length 521;

Best Local Similarity 28.2%; Pred. No. 1.1e-16;

Matches 105; Conservative 50; Mismatches 127; Indels 90; Gaps 15;

QY 84 GRAVNGSQLGKDYIQLSKLLPIRIYSRASLYGPNIGRKNVIALLDGF-----M 134

Db 186 GASVSAELYGKDLIVLKDVIN--KVYKNSWLHKPLVAP-----GGFYBQQWYTKLL 235

QY 135 KVAG-STVDVAVTWHQCYIDGR-----VVKVMDFLKTRLLDLSQIRKIQKVVNTYTPG 187

Db 236 EISGSPVVDVVT--HHIYNLGSGNDPALVKKIMD---PSTLSQVSKTFKDVNTIQEHGP- 290

QY 188 KKWLEGVVTTASG-----GTNNLSDSYAAGFLWNLNTLGLMANGQIDVIRHFFDHGYNH 243

Db 291 ---WASPWGESGAYNSGGRHVSDFIDFSWYLDQLGMSARHNTKVYCRQTLVGGFYGL 347

QY 244 LVQDNFPLPDYWLSSLYKRLGPKVLAVHVGAGLRKPRGRVIRDKLRIYAHCTNHHNH 303

Db 348 LEKGTFPNPDPYYSALLIHLRLMGKGLAVQTDG-----PPQLRVYAHCSKG--- 393

QY 304 NYVGGSTLFLINLH-----RSRKKIKLAGTLRDKLV----- 335

Db 394 ---RAGVTLLILNLSQSDFTVSVSNGINVVVLAESRKKKSLDITLKRFSWTGSKASDG 450

QY 336 ----HOYLLQPYQEG-LKSKSVQLNGQPLVWVDDGTLPKLPRLRAGRTLVIPVPTWG 390

Db 451 YLNREEVHLTP--ENGVLRSKTVNLGCKSLKPTATGDIPLSLPVLRSVNSPLNVLPLSMS 508

QY 391 FVVKVNNALAC 402

Db 509 FIVLPNFDASAC 520

RESULT 3

T01953

hypothetical protein T2L5.6 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 21-Jan-2000

C:Accession: T01953

R:Geisel, C.; Smith, A.; Le, T.

submitted to the EMBL Data Library, October 1998

A:Description: The sequence of A. thaliana T2L5.

A:Reference number: 214470

A:Accession: T01953

A:Status: translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-190 <GET>

A:Cross-references: EMBL:AF096371; NID:g3695386; PID:g3695392

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 4

A:Introns: 36/2; 69/3

A:Note: T2L5.6

C:Superfamily: Arabidopsis thaliana hypothetical protein T2L5.6

Query Match 7.3%; Score 155; DB 2; Length 190;

Best Local Similarity 24.9%; Pred. No. 2.5e-05;

Matches 49; Conservative 32; Mismatches 70; Indels 46; Gaps 6;

QY 233 RSHFFDHGYNHLVDQNFNPLPDYWLSSLYKRLGPKVLAVHVGAGLRKPRGRVIRDKLR 292

Db 12 RQSLIGNGYGLLNTFTNPNPDYYSALINRQLMGRALFTTFSGTK-----KIR 60

QY 293 IYAHCTNHHNHNVRGSGITLFIINLHRSR---KKIKLAGTLRDKLVHQYLLQPY----- 343

Db 61 SYTHCARQSK-----GITVLLMNLDTTIVAKVELNLSF--SLRHTKHKMSYKRASSQ 112

QY 344 ---QEG-----LKSKSVQLNGQPLVWVDDGTLPKLPRLRAGRTLVIP 385

Db 113 LFGGPNGVIOREEVHLTAQDGNLHSQTMLLNGALQVNSMGDLPPIEPIHINSTEPITIA 172

QY 386 PVTMGFFVVKVNNALAC 402

Db 173 PYSIVFVHMRNVVVPAC 189

RESULT 4

T14022

reverse transcriptase homolog - slime mold (Dictyostelium discoideum) retrotransposab

C:Species: Dictyostelium discoideum

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000

C:Accession: T14022

R:Winckler, T.; Tschepe, C.; de Hostos, E.L.; Jendretzke, A.; Dingermann, T.

Mol. Gen. Genet. 257: 655-661, 1998

A:Title: Tdd-3, a transfer RNA gene-associated poly(A) retrotransposon from Dictyoste

A:Reference number: 217958; MUID:98265925

A:Accession: T14022

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-1260 <WIN>

A:Cross-references: EMBL:AF002669; NID:g2558820; PID:g2558822; PIDN:AAC48324.1

Query Match 4.8%; Score 102.5; DB 2; Length 1260;

Best Local Similarity 20.3%; Pred. No. 7.9;

Matches 86; Conservative 72; Mismatches 160; Indels 105; Gaps 22;

QY 65 SNTYSNLILTEPNNYRTMHGRAVNGSQLGKDYIQLSKLLQPIR-IYSRASLYGPNIGRPR 123

Db 304 SSILSIMLKNTDFLLNGYSRELSSNHSISFSQLNSLLNKIKLYTEFOKQNDYNNKAN 363

QY 124 -KNVIALLD-GPMKVAGSTVDVAVTW-----QHC-----YIDGRVVK-VMDF 161

Db 364 IKNLISLLETFEKDQAFATLSAINESKREQLKOELNNYCEETSLKYISARIKKRHDF 423

QY 162 LKTRLLDLSQIRKIQKVVNTYT-----PGKKILEGVVVTTSAGGTNNLS 207

Db 424 TINAVKDTGRTINKQELIEEYVVKYSNLYDKEDDPPSHYEILLENWTVTRDSTWDLNLE 483

QY 208 DSYAAGFLW--LNTGLMLANQIDVIVIRHSFFDH-----GVN-----HLVDQNF 249

Db 484 NEETSOEILEVIKQLNPHKSPGPDGIPNLFYITHREKAPILASAFNDLRLNPHLISKNY 543

QY 250 -----NPLPDYWLSSLYK---RLIGPKVLAVHVGAGLRKPR---PGR 285

Db 544 KEGLIITIPKGDPELIKRRRPTITLANCIYKIHSHKLNINRILPILTKVINHNKQGFVGR 603

QY 286 VIRDKLRIYAHCTNHHNHNVRGSGITLFIINLHRSRKKIKLAGTLRDKL--VHQYLLQPY 343

Db 604 FILHNIISNELINYCNDRKINGIITLYI-----SKLLTRSTVQSQTITTHQHSNOYI 658

QY	344	GO-EGLSKS---	VOLNGQPLVMVDG	TLP-ELKPRPLRAGRTLVTPPYTMGEFVVVKVN	398
		: :	: : :	: :	:
Db	659	NLIINLTKSEARIE	NGRT-----	TIPFEIK-RGVKQG----	DPLSPTLFLVLV-IE 704
QY	399	ALA 401			
Db	705	ALA 707			

RESULT 5
S58740
cytochrome-c oxidase (EC 1.9.3.1) chain I - yeast (Hansenula wingei) mitochondrion
C;Species: mitochondrion Hansenula wingei
C;Date: 19-Mar-1997 #sequence_revision 09-May-1997 #text_change 07-Dec-1999
C;Accession: S58740
R;Sekito, T.; Okamoto, K.; Kitano, H.; Yoshida, K.
Curr. Genet. 28, 39-53, 1995
A;Title: The complete mitochondrial DNA sequence of Hansenula wingei reveals new characters
A;Reference number: S58740; MUID:96022424
A;Accession: S58740
A;Molecule type: DNA
A;Residues: 1-535 <SEK>
A;Cross-references: EMBL:D31785
A;Note: the authors translated the codon CAA for residue 54 as Gly
C;Genetics:

Query Match	4.7%	Score 101;	DB 2;	Length 535;
Best Local Similarity	22.2%	Pred. N6.3;		
Matches	68;	Conservative 46;	Mismatches 102;	Indels 90; Gaps 19;
Qy	106	IRIYSRASLYCPNIGRPKKNVIALLDGPMKVAGSTVDATVW-QHCYIDGRVVKVWDFLKT	164	
Db	260	VSTYSKKPVFGE-----ISMVYANASIA--FLGLFVWSHMYIYG-----LD-ADT	302	
Qy	165	RLDLTSDQIRKKIQKVVNTYTPGKKIWLEGVVTYTSAGGTTNLSDS--YAAGFLWLNTLG-	221	
Db	303	RAYTSTWTVATVGTGKIFS-----WL-----ATLYGGSIRLAVPMLYAI AFLFLTIGG	353	
Qy	222	-----MLANQGDVIVRHSPFDGHYNHLV-----DQNFNPFLPDYW-----LSLYKR-----	263	
Db	354	LTGVALANASLDVAFHDTYVYVVGHFHYVLSMGAIFSLFAGYVYWSQPILGLYFNERLAQI	413	
Qy	264	-----LTGPKV--LAVHVAGLQRPRGPRVTRDKLRIYAHCTNHHNHNHYVG-----	308	
Db	414	QFWLIFVGANYFPMHFGLQGMPR-----RIPDYPDAYAGWNYVSSIGSVIAII	464	
Qy	309	SITLFIINLHRSRKKIKLAGTLRDLKVHO---YLLQP-----YGEGLKSKSVQ--L	355	
Db	465	SLALFYIIYD-----QLINGLTNKIDKNKSVVSKAPDFVESNTIFANNKSASIEFL	519	
Qy	356	NGQPLV	361	
Db	520	NSPPAI	525	

RESULT
C71374

probable glucose inhibited division protein A (gidA) - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 26-Aug-1999
C:Accession: C71374
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dod-
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterb-
rath, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-386, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spiro-
A:Reference number: AT1250; MUID:98332770
A:Accession: C71374
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-630 <COL>
A:Cross-references: GB:AE001189; GB:AE000520; NID:3322293; PIDN:AAAC65038.1
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0044
A:Superfamily: gidA protein

Query Match	4.7%	Score 100.5;	DB 2;	Length 630;
Best Local Similarity	21.3%	Pred. No. 4.2;		
Matches	77;	Conservative	52;	Mismatches 129; Indels 103; Gaps

Qy	89	GSQLGK--DYIQKSLILQPIRIYSRASLYGPNIGRPRKKNVIALLDGFMKVAGSTVDVAVTW	146
Db	68	GGWGKFDADCMIQ-----YRLNKSr--GPVQAPRIQADKEL--YAQVKVYILECTQH	118
Qy	147	QHCVIDGRVVKVMDFLKTRLLDLSQIRKIQKVVNTYTPGKKIWLEGVVTTSAGCTNNL	206
Db	119	LHLQYD---TVVDVCSNTDAGYVAYGAHAVVT--ARGRISARAVVLTG-----	166
Qy	207	SDSYAAGFLWL-----NTGLMLANGIDVIRHGSFFDHG-----YNHLVDQNF	249
Db	167	--TMEGRVYIGEYEAPEGRGLGEHAEGLAALKRKKGFGOMGLTKGTPARVLRSVD---	221
Qy	250	NPLPDYWLSLYKR---LIGPKVLAVHVAGLQKPRGVRIRDKLRIYAHCTNNHHNH	305
Db	222	-----LSVMEKQDAIMRPFSA-HV-----EINRPHADCYINYTNERTHQL	263
Qy	306	VRGSITLFIINLHRS---RKIKLAGT-----LRDKLVHQYLLQPYGQGLKS	350
Db	264	IRE-----NFRHSFSGRIKAVGTRYCPSTEDKVKRFPDRIRHOLYIEP---EGLDT	313
Qy	351	KSVOLNGOPLVWVDD-----GTLPELK-----PRPLRAGRTLVPVPMGF-FVWKNVA	399
Db	314	EELYINGLSSCLPEDIQDEMIRTPGHERAVITRPVAVDYAVLFPVQIGIDLOTRVSG	373
Qy	400	L 400	
Db	374	L 374	

RESULT	7
E84992	
trypophan--trna ligase (EC 6.1.1.2) [imported] - Buchnera sp. (strain APS	
N:Alphabet names: trypophanyl-trna synthetase	
C:Species: Buchnera sp.	
C:Accession: E84992	
R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.	
Nature 407, 81-86, 2000	
A:Title: Genome sequence of the endocellular bacterial symbiont of aphids	
A:Reference number: A84930; MUID:20445173	
A:Accession: E84992	
A:Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-335 <STO>	
A:Cross-references: GB:AP000398; GSPDB:GN00144	
A:Experimental source: strain APS	
C:Genetics:	
A:Gene: trps; BU536	

QY 341 QPYGEGELKSKSVOLNGQPLVMVDGSL-LPELKPRLR--AGRTIVIPVTMGFFV 393
Db 413 QTYDYNKNGNWAFSDWVGDLFLNNTTHRALRSLANPTLYDQPDNNNTV 468

RESULT 13
T51619
probable ethylene receptor [imported] - netted muskmelon
C:Species: Cucumis melo var. reticulatus (netted muskmelon)
C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 08-Dec-2000
C:Accession: T51619
R:Sato-Nara, K.; Yuhashi, K.I.; Higashi, K.; Hosoya, K.; Kubota, M.; Ezura, H.
Plant Physiol. 120, 321-330, 1999
A:Title: Stage- and tissue-specific expression of ethylene receptor homolog genes during
A:Reference number: 225417; MUID: 99252277
A:Accession: T51619
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-740 <SAT>
A:Cross-references: EMBL:AF054806; PIDN:AAC99645.1
C:Genetics:
A:Gene: ETR1
C:Superfamily: response regulator homology

Query Match 4.5%; Score 95; DB 2; Length 740;
Best Local Similarity 18.9%; Pred. No. 15;
Matches 87; Conservative 79; Mismatches 143; Indels 152; Gaps 23;

QY 20 DQIVRSVALD---KQKCKIAHQPDVMLELQREKAAQMHLY-----LKEQFSNTYSNL 71
Db 321 DLLMEQNALDLARAEATAIRANDFLAVMNHMRTPMHAIALLSSLLQETELTPEQRL 380

QY 72 ----ILTEPNNTYTHMGRVAVGSLQKDYIQLK-----SLQPIRYSRAS 113
Db 381 MVETILKSSNLLATLNDVLDLSRLDGLDGLDIGNLHAFVEKVLNLIKPVTLVKKLS 440

QY 114 L---YGNP-----IGPRKNVIALI-----DGFMKVAGSTVDVAVTQHCVIDGR 154
Db 441 LTLHGLPDLVPFVAGDEKRLMAQLNVGNVAVKFSKESGISAIYAKSETF-----R 493

QY 155 VVKVMDP-----LKRLLDRLS-----DQIRKIQKVNTYTPGKKIWLEGVVTTSA 200
Db 494 EIRVPDFHPVPSDSHYLRVQVQKGTSGISPDIPKLTFFKAQTVGPR-----NS 544

QY 201 GGTN-----NLSDSYAGFLWNTLGLMLANGIDVIRHSFFDGHYHLYDQNFN 250
Db 545 GSGGLGLATCKRFVNLME---GHWLESEGLGKGTATFIVKLGTD-----QSNES 593

QY 251 PLPDYWLSLLYKRLI-----GPKVLAVHVGLOKRPGRVIRDKLRIYAHCTNHHNNY 305
Db 594 KLP--YTSKIHENSHTSPFGLKVLVMDNGVSRVTKG-----LLVHLGC----- 637

QY 306 VRGSITFLTIINLHRSRKKIKLAGTLRD-----KLHVQYLLQPYGQEG-----LKS 350
Db 638 -----EVTAGSIEEFLRVVVSQEHKVVFMFDICTP-GVDGYELAIRIRE 679

QY 351 KSVQLNGQPLVMVDGTLPEL-KPRLRAGR-LVIPPVTM 389
Db 680 KFAKCHERPVMVLTGNSDKVTKESCLRAGMDGLILKPVSI 720

RESULT 14
G90554
p46-like (mycoplasma hyorhinis) lipoprotein [imported] - Mycoplasma pulmonis (strain UAE
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: G90554
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A:Reference number: A99512; MUID: 21267165; PMID: 11353084

A:Accession: G90554
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-460 <KUR>
A:Cross-references: GB:AL445566; PID:gl14089757; PIDN:CAC13516.1; GSPDB:GN00153
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYPUL_3430
A:Genetic code: SGC3

Query Match 4.4%; Score 94.5; DB 2; Length 460;
Best Local Similarity 19.6%; Pred. No. 8.3;
Matches 78; Conservative 53; Mismatches 145; Indels 121; Gaps 18;

QY 60 LKEQFSNTYSNLLTEPNNTYTHMGRVAVGSLQKDYIQLKSLQPIRYSRASLYGPN 119
Db 55 VSEKISDGSIVITVDPENQR-----WVETKKQLD-----AYSKT----- 89

QY 120 GRPRKNVIALLDGFMKVGAGSTVDVAVTQHCYIDGRVVKVMDFLKTRLLDLSQ----- 173
Db 90 -----NGFEHSSNHVKAQEQNSFVDAELAKTGNNKPKVVLGMAADSGNATQA 138

QY 174 -----IRKIQK-----VVNTYTPGKKIWLEGVVTTSA-----GQTNNL 206
Db 139 IBSTTNAQAQFTAVDRFTHKISNNYNWYVAPNNYTVGQ---LQGLALISGIYKGQGEPEK 195

QY 207 SDSYAAGFLWNTL-----GMLANQGI-DVVIHRSFEDHGYNHL-----VDQNFNPLPDY 256
Db 196 TLEEAKTYVMANKLASEKGFVALAGAPEDNNSHLFEKGAAMDVITATMKIDSNLKYFGE 255

QY 257 LSLLYKRLIGPKVLAVHVGLOKRPGRV---IRDKLRIYAHCTNHHNNVVRGS---I 310
Db 256 GADVFNRLSG-----DSKPTDQELTSAITSHFRVVAQA---NWNYSQKGQFLD 300

QY 311 TLFIINLHRSRKKIKLAGTL--RDKLHVQYLLQPYGQEGLSKSVOLNGQPLVMVDGTL 368
Db 301 TLFKASTFDSRKN-NLAAVLAPNDEMAQAAITSEQKGLDKPKIVITGQ-----DSNQ 353

QY 369 PELKPRPLRAGRTLVP-----PVTMGFFVKN 396
Db 354 PSLSRIDNETGQNMFTISKEDWKIAAISTALAYIVKN 390

RESULT 15
G86855
glucose inhibited division protein Gida [imported] - Lactococcus lactis subsp. lactis
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: G86855
R:Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Eh
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
A:Reference number: A86625; MUID: 21235186; PMID: 11337471
A:Accession: G86855
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-625 <STO>
A:Cross-references: GB:AE005176; PID:gl12724877; PIDN:AAK05945.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: gida
C:Superfamily: gida protein

Query Match 4.4%; Score 94; DB 2; Length 625;
Best Local Similarity 19.6%; Pred. No. 14;
Matches 72; Conservative 58; Mismatches 119; Indels 118; Gaps 17;

QY 100 KSLLQPIRYSRASLYGPNIGRPRKNVIAL-LDGFMKVAGSTVDVAVTQHCYID---GR 154
Db 31 KYTLMTINLWVAFMPCNPISGSAKGIYVREIDALGGEMGRNIDKTYIQMKMLNTGKGP 90


```

Qy 155 VVKVM-----DFLKYRLDLTSDQ-----IRKI-----QKVVNTYT-PGKKIWL 192
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 91 AVRALRAQADKDEVADSMKNTVSQENLTLROGMVEELILDEKKKVIKGTCTKYGA 150

Qy 193 EGVVTT-----SAGCTNNLSDSYAAGFLWLNLGLMLANQGDVIVRHSF 236
   | : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 151 KAVIITGTALRGEIITIGELKYSSGPNNSLS-----SIGLADN-----LREIG 193

Qy 237 FDHCYNHLVDQENPLDPYWLSLLYKRLICPKVLAVHVAGLQKRPFG----- 284
   | : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 194 FEIG-----RFTGTPRVLASSIDYDTEIQGDGDEAPNHFSPWSS 234

Qy 285 --RVIRDKLRIYACHTNNHNNVVRGSIITLFINLHRS---RKKIKLAG-----TLROKL 334
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 235 DENVLKDOIICWLTYYTENSHTILRG-----NLHRAPLFGSVKVGVPKPCPSIEDKI 287

Qy 335 V-----HOYLLQPYGQEGSKSKSVQLNGQPLVMVWDGTLPELKPRLRAGRTVLPIPV 387
   | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 288 TRFADKPRHOLFLEP---EGRNTEEVYIGGLSTSMPEDEVQFDVLKVSIPCLENAQMRRFY 344

Qy 388 TMGFFVV 394
   : : | : |
Db 345 AIEYDVV 351

```

Search completed: July 30, 2002, 08:16:32
Job time: 349 sec

THIS PAGE BLANK (USPTO)

QY	147	QHCYIDGRVVKVMDFLKTRLLDTLSQIRKIQKVVNTYTPGKKIWLGVVTTSGAGTNNL	206
DB	119	LHLQYD-----FVVDVCSNTTDAAGVYGAHAHVVT--ARGRRISARAVLTTG-----	166
QY	207	SDSYAAGFLWL-----NTGLMANQGDVIVIRHSFDDHG-----YNHLVDQNF	249
DB	167	--TFMEGRVYIGEYBAPEGRLGEHAAEGALRRKKGFQMLTKGTPTARVLRKSV-----	221
QY	250	NLPDYWLSLLYKR-----LIGPKVLAVHAGLQRPGRVIRDKLRIYAHCTNNHHNHY	305
DB	222	-----LSVMEKQADAIMRPFSA-HV-----EINRHADCYINVTNERHQH	263
QY	306	VRGSITLFTINIHR-----RKKTKLAGT-----LRDLVHQVLLQPYQOEGGLKS	350
DB	264	IRE-----NHRSPFSGRIKAVGTGYPCPSIEDKVRKFPDRIRHQYIEP-----	313
QY	351	KSVALNQPLVMDV-----GTLPELK-----PRPLRAGRLVLPVPTMGF-FVVKVNA	399
DB	314	EELYINGLSSCLPEDIQDEMIRTFQMERAVITRPAYADVAVLPVQVGLDQLTKRVSG	373
QY	400	L 400	
DB	374	L 374	
RESULT 2			
ID	SYW_BUCAI	STANDARD;	PRT; 335 AA.
AC	P57602;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DE	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	tryptophanyl-tRNA synthetase (EC 6.1.1.2) (tryptophan--tRNA ligase)		
DE	(TrpRS).		
GN	TRPS OR BU536.		
OS	Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum		
OS	symbolic bacterium)		
OC	Bacteria; Proteobacteria; gamma subdivision; Buchnera.		
OX	NCBI_TaxID=118099;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=TOKYO 1998;		
RX	MEDLINE=20445173; PubMed=10993077;		
RA	Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;		
RT	genome sequence of the endocellular bacterial symbiont of aphids		
RL	Buchnera sp. AFS.;		
RL	Nature 407:81-86(2000).		
CC	-1- CATALYTIC ACTIVITY: ATP + L-tryptophan + tRNA(Trp) = AMP +		
CC	diphosphate + L-tryptophanyl-tRNA(Trp).		
CC	-1- SUBUNIT: HOMODIMER (BY SIMILARITY).		
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic.		
CC	-1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; AP001119; BAB13229.1; -		
DR	InterPro; IPR001412; TRNA-synt_1.		
DR	InterPro; IPR002306; TRNA-synt_trp.		
DR	Pfam; PF00579; TRNA-synt_lb.1.		
DR	PRINTS; PR01039; TRNASYNTHTRP.		
DR	PROSITE; PS00178; AA_TRNA_LIGASE_I.1.		
KW	Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;		
KW	Complete proteome.		
FT	SITE	14 22	"HIGH" REGION.
FT	SITE	196 200	"KMSK" REGION.
FT	BINDING	199 199	ATP (BY SIMILARITY).
FT	SEQUENCE	335 AA; 38549 MW; 871562D5A6734E3F	CRC64;

147	QY	QHCVIDGRVVKVMDFLKTRLLDTLSQIRKIQKVVNTYTPGKKIWLEGVVVTSAGTNNL	206
148	QY	: :	

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: L12148; AAA16444.1; -;
 InterPro: IPR001343; HemLysn_Ca_bind.
 InterPro: IPR003355; RTX_N.
 InterPro: IPR003995; RTX_A.
 Pfam: PF00353; hemolysinCabinid; 1.
 Pfam: PF02382; RTX; 1.
 PRINTS: PR00313; CABNDNGRPT.
 PRINTS: PR01488; RTXTOXINA.
 PROSITE: PS00330; HEMOLYSIN_CALCIIUM; 4.
 Hemolysis; Toxin; Cytotoxicin; Repeat; Calcium;
 Transmembrane; Lipoprotein; Palmitate.
 TRANSMEM 154 170 POTENTIAL.
 TRANSMEM 312 333 POTENTIAL.
 TRANSMEM 393 414 POTENTIAL.
 DOMAIN 625 780 7 X REPEATS, GLY-RICH.
 REPEAT 625 630 1.
 REPEAT 730 735 2.
 REPEAT 739 744 3.
 REPEAT 748 753 4.
 REPEAT 757 762 5.
 REPEAT 766 771 6.
 REPEAT 775 780 7.
 SEQUENCE 947 AA; 101559 MW; 9744F06395EF5BED CRC64;

Query Match 4.4%; Score 93.5; DB 1; Length 947;
 Best Local Similarity 24.3%; Pred. No. 9.3;
 Matches 64; Conservative 30; Mismatches 80; Indels 89; Gaps 13;

QY 13 YLKNVEDIVRSVALDKQCKTAQHPDVMLELQREK-----AAQHVLVLLKEQFSNT 67
 Db 43 YPKDYEDVSGRGGLQDLVRAE-----DLGIEVOREERNGIATAOANSLSTIONILGFS 97
 QY 68 YSNLILTEPNNTYTHMGRVAVNGSQLGKDYIOLKSLLOPIRIYVSRLSYGPNIGRPNKVI 127
 Db 98 ERGVVLSAP-----QLDKLLQYKI-SKAPGSSENVAKNLGNAQ 135

QY 128 ALLDGFMKVAGSTVDVAVTQHCYIDGRVVK-----VMDFLKTRL-----L 167
 Db 136 TLLSGIQSLGSMAGMD-----LD-EILKNKGSELDLAKAGLELTNSLIENIANSVQTL 189

QY 168 DTLSDOI-----RKIQVNTYTPGKK-----TWLEGVVTSAGTNN--LSDSY 210
 Db 190 DTFSQISQLGTLQNVKGLGLTGDKNFSGFKAGLGLEIVSGLLSGATAALVLADKN 249

QY 211 A-----AGELWLNTLGLMLAQ 226
 Db 250 ASTDRKVGAGF-----ELANQ 265

RESULT 6
 SNK_RAT
 ID SNK_RAT STANDARD; PRT; 682 AA.
 AC Q9R012;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE serine/threonine-protein kinase SNK (EC 2.7.1.-) (Serum inducible kinase).
 DE SNK.
 GN SNK.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99452760; PubMed=10523297;

RA Kauselmann G., Weller M., Wulff P., Jessberger S., Konietzko U.,
 RA Scafidi J., Staubli U., Bereiter-Hahn J., Streibhardt K., Kuhl D.;
 RT "The polo-like protein kinases Fnk and Snk associate with a Ca(2+)-and
 RT integrin-binding protein and are regulated dynamically with synaptic
 RT plasticity";
 RL EMBO J. 18:5528-5539(1999).
 CC CC FUNCTION: MAY PLAY A ROLE IN THE DIVISION OF AT LEAST SOME CELL
 CC TYPES, SUCH AS FIBROBLASTS, AND COULD FUNCTION IN EMBRYOGENESIS,
 CC WOUND HEALING OR NEOPLASIA (BY SIMILARITY).
 CC -I- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC CDC5/POLO SUBFAMILY.
 CC -I- SIMILARITY: CONTAINS 2 POLO BOX DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL: AF136583; AAF08366.1; -;
 HSP: P00518; lphk
 InterPro: IPR000719; Euk_pkinase.
 InterPro: IPR000959; POLO_box.
 InterPro: IPR002290; Ser_thr_pkinase.
 Pfam: PF00069; pkinase; 1.
 Pfam: PF00659; POLO_box; 2.
 SMART: SM00220; S_TKC; 1.
 DR PROSITE: PS00078; POLO_BOX; 2.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 KW Transferase; Serine/threonine-protein kinase; ATP-binding; Repeat.
 FT DOMAIN 54 59 POLY-HIS.
 FT DOMAIN 79 331 PROTEIN KINASE.
 FT NP_BIND 85 93 ATP (BY SIMILARITY).
 FT BINDING 108 108 ATP (BY SIMILARITY).
 FT ACT_SITE 202 202 BY SIMILARITY.
 FT DOMAIN 507 570 POLO BOX 1.
 FT DOMAIN 603 674 POLO BOX 2.
 SQ SEQUENCE 682 AA; 77919 MW; 58C50DEBDE83DF3 CRC64;

Query Match 4.3%; Score 92.5; DB 1; Length 682;
 Best Local Similarity 22.7%; Pred. No. 7.2;
 Matches 94; Conservative 58; Mismatches 161; Indels 101; Gaps 21;

QY 6 RGGPGDYLYLKNVEDIVRSVALDKQCKTAQHPDVMLELQREKAAQ-----M 55
 Db 87 KGG-----FAKCYE-----MTDTNNKYAAKLIHRSVAKPHQREKIDKEIHLHLHK 137

QY 56 HLVLKKEQFSNTYSLILTEPNNTYTHMGRVAVNGSQLGKDYIOLKSLLOP-IRIYSRASL 114
 Db 138 HVQFYHYFEDKENIYILLEYCSRSM-----AHILK-----ARKVLTEPEVRYLRQIV 187

QY 115 YGPNIGRPNKVNIALDGFMKVAGSTVDVAVTQHCYIDGRVVKVMDFLKTRLDTLSDOI 174
 Db 188 SGLKYLHEQE-----ILHRDLKLGNNFIN-----EAMELKVGDFGLAARLEPLEHRR 234

QY 175 RKIQVNTYTP-----GKKIWLEGVV--TTSAG-----GTNLSDSY-----AAGF 214
 Db 235 RTICCTPNVLSPEVUNKOGHGESDIWALGCVMTLGRPPFTTNLKYRCIARY 294

QY 215 LWLNTL-----GMLAQ-----GIDVIRHRSFFDHGY--NHLVDQNFNPLDPDWLS 258
 Db 295 TMPSSLAPAKHLIASMLSKNPEDRPSLDDIIRHDFLQGFDPDRLSSCCCTVDPFHLS 354

QY 259 LLYKLLGPKVLAVHVGQRPR-----PGRVIRDLRIYACTNHHNHHNHYRGSTTLFI 314
 Db 355 SPAKNFF-KAAAAALFGGKKKARYNDTHNKVSKDEDIY-----KLRLDKAKTSITQ-Q 407

QY 315 INLHRSRKIKLAGTLRDLKVLHQLLPQYQEGSLKSKSVQLNGQPLVMVDDGTL 368

SQ SEQUENCE 1616 AA; 188063 MW; F17D7C452D123F5 CRC64;

Query Match 4.3%; Score 92.5; DB 1; Length 1616;
Best Local Similarity 19.7%, Pred: No. 23;
Matches 83; Conservative 69; Mismatches 141; Indels 129; Gaps 19;

QY 17 NYEDDIVRSDVA--LDKQRGCKIAQHPDYML-----ELQREKAQAOMHLVLKK 61
| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 205 NFDKCITSEIEIAGLRYSSCEPECKEKTDLIRPQTVTYVILENEELEKESEVRSLTYNVN 264

QY 62 EQ---PSNTYSNILTEPNNYRTMHGRAVSOLGKDYYQLKSLLQPRIYRSASLYGN 118
| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 265 GOEVMKETRSKLVEENHSIKS-HYEKYNGE---KESI-----ITS----- 302

QY 119 IGRPRKNVIALLDGFMKVAGSTVDATVWTOHCYIDGRVVVKVMDFLTRLDTLSDDIRKI 178
| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 303 -----SRWEQLVEDFEK-NGDKAEFAFFPKPLD-----KKMHLIKTITEIQEVE 347

QY 179 KVNWTYPGGKINLEGV----TTSAGGTNNISDSYAAGFLWLNLGLMANOGIDVIRH 234
| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 348 NN-----PETSHFLARLVIRFTTS---TSQLKEIH-----ETLYVKADKKIOSLMEH 393

QY 235 SF-----FDCHYNHLVDQNFPDPYWLSSLYKRRLIGPKVLAVHVHVGALQRPGRV 286
| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 394 ALATAGTKNTYIOHTLVHIENEDIVPLE-----AAQLLKSIQETPPSPQT 437

QY 287 IRDKLRIVAHCTNHNNHYNR-----GSITFIILNHRSR----- 321
| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 438 IAELAIIKFAESRVSKNQVYROSALGAAGSVRGVINDYKNIRPLVREDRKLEKFLRFV 497

QY 322 -KTKTLAGTRDKLVHOYLLOPYGOEQGLSKSKSQVLPMVDDTGTPLPKPRLRAGR 380
| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 498 MQOYKDAETTEKIL---ALKSIGNAGLDISVNOINE---IIVDKQPLLVPKREAIDLAR 551

QY 381 TL 382
|
Db 552 LL 553

RESULT 9
TEGU_HCMVA STANDARD; PRT; 2241 AA.
AC P16785;
DF 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Probable large tegument protein.
GN HAd69.
OS Human cytomegalovirus (strain Adl69).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90269039; PubMed=2161319;
RA Chou S.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,
RA Horanell T., Hutchison C.A. III, Kouzarides T., Martignetti J.A.,
RA Readle E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;
RT Analysis of the protein-coding content of the sequence of human
cytomegalovirus strain Adl69."
RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).

CC -! FUNCTION: TEGUMENT PROTEIN.
CC -! SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC EBV-1 24, EBV BPFL1, HVIS-1 64, VZV 22, AND HCMV UL48.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licensese@isb-sib.ch).

Search completed: July 30, 2002, 08:31:33
Job time: 980 sec

```
RESULT 15
UVRE_PASMU
ID UVRE_PASMU STANDARD; PRT; 678 AA.
AC P57844;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Excinuclease ABC subunit B.
GN UVRB OR PM0429.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRB STIMULATES
CC THE ATPASE ACTIVITY OF UVRA IN THE PRESENCE OF UV-IRRADIATED
CC DOUBLE-STRANDED DNA. IT ALSO ENHANCES THE ABILITY OF UVRA TO BIND
CC TO UV-IRRADIATED DUPLEX DNA (BY SIMILARITY).
CC -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE UVRB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE006079; AAK02513.1; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR001943; UVR.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF02151; UVR; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR SOS response; Excision nuclease; DNA repair; ATP-binding;
KW Complete proteome.
FT NP_BIND 44 51
FT SEQUENCE 678 AA; 77401 MW; 2EE928B9A23ADF15 CRC64;
Qy Query Match 4.2%; Score 89; DB 1; Length 678;
Best Local Similarity 21.1%; Pred. No. 14;
Matches 40; Conservative 38; Mismatches 82; Indels 30; Gaps 6;
Qy 9 PCPDYLLKNYE-----DDIVRSVALDKQKCK-IAQHPDVMLELQREKAAQMHLV 58
Db 401 PGP-----YELEKSGGFIIDQVVRPTGLDPEIRPVAIQVDDLLSEARQADQNERV 454
Qy 59 LIKEQFSNTYSNLIITPENNYRTMHGRAVNGSQLGKDYIQLKSLLOPIRYSRASLYGPN 118
Db 455 LV-----TTLTKKMAEDTDLDEHGIRVRYLHSDIDTVERVEIIRDLRGEFDVLVGIN 509
Qy 119 IGR-----PRKNVIALLD-----GFMKVAGSTVDVATWQHCHYIDGRVVKVMDFLKRLTLD 169
Db 510 LLREGLDIPEVSLVAILDADKEGFLRSERSLIQTIGRAARNLKGKAILYADRTINSQKA 569
Qy 170 LSDQIRKIQK 179
Db 570 IETNRREK 579
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:30:47 ; Search time 123.94 Seconds
(without alignments)
565.297 Million cell updates/sec

Title: US-09-836-461-2_COPY_130_534
Perfect score: 2129
Sequence: 1 NPAKSRGGPGDYLYKNYED.....PVTMGFFVKNVNALACRYR 405

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2129	100.0	534	4 Q9HB38	Q9hb38 homo sapien
2	2090	98.2	592	4 Q9HB37	Q9hb37 homo sapien
3	1822	85.6	480	4 Q9HB39	Q9hb39 homo sapien
4	810.5	38.1	536	11 Q9QZF8	Q9qzf8 rattus norv
5	795.5	37.4	543	4 Q9V251	Q9v251 homo sapien
6	795.5	37.4	545	4 Q9UL39	Q9ul39 homo sapien
7	785	36.9	545	6 Q9MY0	Q9my0 bos taurus
8	736.5	34.6	523	13 Q9QYK5	Q9qyk5 gallus gall
9	301	14.1	521	10 Q9SDA1	Q9sda1 arabidopsis
10	301	14.1	543	10 Q9FF10	Q9ff10 arabidopsis
11	258.5	12.1	516	10 Q9FLK8	Q9flk8 arabidopsis
12	236	11.1	536	10 Q9FZP1	Q9fzpl arabidopsis
13	215.5	10.1	527	10 Q9LRC8	Q9lrc8 scutellaria
14	155	7.3	190	10 Q82604	Q82604 arabidopsis
15	155	7.3	935	5 Q9VE79	Q9ve79 drosophila
16	110	5.2	174	10 Q9ATW5	Q9atw5 zea mays (m

17	103	4.8	768	3	Q9P8N9	Q9p8n9 cladosporiu
18	102.5	4.8	1260	5	O15639	O15639 dictyosteli
19	102	4.8	885	5	O61232	O61232 lymanaea sta
20	98	4.6	491	2	O08457	O08457 clostridium
21	98	4.6	1090	12	O91E95	O91e95 human rotav
22	97	4.6	1020	5	O25415	O25415 leishmania
23	97	4.6	4533	5	O9BIX3	O9bix3 tetrahymena
24	96.5	4.5	1429	2	O9LAP7	O9lap7 alteromonas
25	96	4.5	898	10	Q9SIB9	Q9sib9 arabidopsis
26	95.5	4.5	551	10	Q38779	Q38779 avena sativ
27	95.5	4.5	663	8	Q9MJ69	Q9mj69 physarum po
28	95.5	4.5	1281	11	O88915	O88915 mus musculus
29	95	4.5	456	5	O22152	O22152 caenorhabdi
30	95	4.5	512	8	O99806	O99806 ixodes hexa
31	95	4.5	557	16	O97G52	O97g52 clostridium
32	95	4.5	740	10	O82436	O82436 cucumis mel
33	94.5	4.4	460	16	O98QL8	O98ql8 mycoplasma
34	94	4.4	855	15	Q90DZ7	Q90dz7 human immun
35	93.5	4.4	1234	5	O9YIH7	O9yih7 dictyosteli
36	93.5	4.4	1219	11	O9QYM3	O9qym3 mus musculus
37	93.5	4.4	1713	9	O94MA1	O94ma1 lactococcus
38	93	4.4	320	17	O97UM5	O97um5 sulfolobus
39	93	4.4	441	10	O94I11	O94i11 arabidopsis
40	93	4.4	503	3	O9HGN1	O9hgn1 schizosacch
41	92.5	4.3	515	10	O49258	O49258 avena sativ
42	92.5	4.3	620	10	O9FUI4	O9fui4 oryza sativ
43	92.5	4.3	685	4	O96CV7	O96cv7 homo sapien
44	92.5	4.3	876	16	O99TH2	O99th2 staphylococ
45	92.5	4.3	1845	12	Q9WJZ8	Q9wjz8 human cytom

ALIGNMENTS

RESULT	ID	PRELIMINARY;	PRT;	534 AA.
Q9HB38	Q9HB38			
AC	Q9HB38;			
DT	01-WAR-2001 (TREMBLrel. 16, Created)			
DT	01-WAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-WAR-2001 (TREMBLrel. 16, Last annotation update)			
DE	HEPARANASE-LIKE PROTEIN HPA2B.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20483645; PubMed=11027606;			
RA	McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,			
RA	Hircock M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.,			
RT	"Cloning and Expression Profiling of Hpa2, a Novel Mammalian			
RT	Heparanase Family Member.;"			
RL	Biochem. Biophys. Res. Commun. 276:1170-1177(2000).			
DR	EMBL; AF282886; AAG23422.1; -.			
SQ	SEQUENCE 534 AA; 60063 MW; C3DE5E90CB338C4 CRC64;			

Query Match 100.0%; Score 2129; DB 4; Length 534;
Best Local Similarity 100.0%; Pred. No. 2.3e-174;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	NPAKSRGGPGDYLYKNYEDIVRSVDALDKQKCKTAQHPDVLMLEQREKAQMHVLVL 60
Db	130	NPAKSRGGPGDYLYKNYEDIVRSVDALDKQKCKTAQHPDVLMLEQREKAQMHVLVL 189
QY	61	KEQFSNTSYNLIITPENNYRTHMGRAVNSQLGKDYQLKSLLOPIRYSRASLYGNIG 120
Db	190	KEQFSNTSYNLIITPENNYRTHMGRAVNSQLGKDYQLKSLLOPIRYSRASLYGNIG 249
QY	121	RPKNVIALDGPMPKVGSTVDATVWQHCVIDGRVVKVMDFLKRLDLTSLDQIRIKQV 180
Db	250	RPKNVIALDGPMPKVGSTVDATVWQHCVIDGRVVKVMDFLKRLDLTSLDQIRIKQV 309

QY 181 VNTYTPGKKIWLGVVTTTSGAGTNNLSDSYAAGFLWNLNTGLMANGQIDVIRHSHFFDHG 240
DB 310 VNTYTPGKKIWLGVVTTTSGAGTNNLSDSYAAGFLWNLNTGLMANGQIDVIRHSHFFDHG 369
QY 241 YNHLVDQNFPLDPYWLSSLYLRLGPKVLAVHVAGLQKRPGRVIRDKLRIYAHCTNH 300
DB 370 YNHLVDQNFPLDPYWLSSLYLRLGPKVLAVHVAGLQKRPGRVIRDKLRIYAHCTNH 429
QY 301 HNNYVGRSTTLFIINLHRSRKKIKLAGTLRDKLVHOYLQPYGQGLKSKSVQLNGQPL 360
DB 430 HNNYVGRSTTLFIINLHRSRKKIKLAGTLRDKLVHOYLQPYGQGLKSKSVQLNGQPL 489
QY 361 VMVDDGTLPKLPRLRAGRTLVIPTMTGFFVYKNNNALACRYR 405
DB 490 VMVDDGTLPKLPRLRAGRTLVIPTMTGFFVYKNNNALACRYR 534
RESULT 2
ID Q9HB37 PRELIMINARY; PRT; 592 AA.
AC Q9HB37;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HEPARANASE-LIKE PROTEIN HPA2C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20483645; PubMed=11027606;
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,
RA Hircok M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;
RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian
RT Heparanase Family Member.";
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).
DR EMBL: AF282887; AAC23423.1; -.
SQ SEQUENCE 592 AA; 66580 MW; 95C384AD9A74258E CRC64;

Query Match 98.2%; Score 2090; DB 4; Length 592;
Best Local Similarity 87.5%; Pred. No. 6.1e-171;
Matches 405; Conservative 0; Mismatches 0; Indels 58; Gaps 1;
QY 1 NPAKSRGGPGDYLYKNYEDDIVRSDVALDKQCKIAQHPDMVLELQREKAAQMHVLL 60
DB 130 NPAKSRGGPGDYLYKNYEDDIVRSDVALDKQCKIAQHPDMVLELQREKAAQMHVLL 189
QY 61 KEQFSNTYSLNLT-----EPNNYRTMHGRAVNGSQLGKDYIQLKSLQPIRYSRASLYGPNIG 74
DB 190 KEQFSNTYSLNLTARSLDKLYNFADCSGLHLIFALNALRRNPNNSSSALSLLKYSA 249
QY 75 -----EPNNYRTMHGRAVNGSQLGKDYIQLKSLQPIRYSRASLYGPNIGTRP 122
DB 250 SKKYNISWELGNEPNNYRTMHGRAVNGSQLGKDYIQLKSLQPIRYSRASLYGPNIGTRP 309
QY 123 RKNVIALDGFEMKVGSTVDATVQHCYIDGRVVKVMDFLKTRLLDLSQIRKIQKV 182
DB 310 RKNVIALDGFEMKVGSTVDATVQHCYIDGRVVKVMDFLKTRLLDLSQIRKIQKV 369
QY 183 TYTPGKKIWLGVVTTTSGAGTNNLSDSYAAGFLWNLNTGLMANGQIDVIRHSHFFDHGYN 242
DB 370 TYTPGKKIWLGVVTTTSGAGTNNLSDSYAAGFLWNLNTGLMANGQIDVIRHSHFFDHGYN 429
QY 243 HVLVDQNFPLDPYWLSSLYLRLGPKVLAVHVAGLQKRPGRVIRDKLRIYAHCTNH 302
DB 430 HVLVDQNFPLDPYWLSSLYLRLGPKVLAVHVAGLQKRPGRVIRDKLRIYAHCTNH 489
QY 303 HNNYVGRSTTLFIINLHRSRKKIKLAGTLRDKLVHOYLQPYGQGLKSKSVQLNGQPL 362
DB 490 HNNYVGRSTTLFIINLHRSRKKIKLAGTLRDKLVHOYLQPYGQGLKSKSVQLNGQPL 549

QY 363 VDDGTLPKLPRLRAGRTLVIPTMTGFFVYKNNNALACRYR 405
DB 550 VDDGTLPKLPRLRAGRTLVIPTMTGFFVYKNNNALACRYR 592
RESULT 3
ID Q9HB39 PRELIMINARY; PRT; 480 AA.
AC Q9HB39;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HEPARANASE-LIKE PROTEIN HPA2A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20483645; PubMed=11027606;
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,
RA Hircok M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;
RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian
RT Heparanase Family Member.";
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).
DR EMBL: AF282885; AAC23421.1; -.
SQ SEQUENCE 480 AA; 53900 MW; F75F89F67AC1FF83 CRC64;
Query Match 85.6%; Score 1822; DB 4; Length 480;
Best Local Similarity 86.7%; Pred. NO. 4.7e-148;
Matches 351; Conservative 0; Mismatches 0; Indels 54; Gaps 1;
QY 1 NPAKSRGGPGDYLYKNYEDDIVRSDVALDKQCKIAQHPDMVLELQREKAAQMHVLL 60
DB 130 NPAKSRGGPGDYLYKNYED-----EPNNYRTMHGRAVNGSQLGKDYIQLKSLQPIRYSRASLYGPNIG 149
QY 61 KEQFSNTYSLNLTPEPNYRTMHGRAVNGSQLGKDYIQLKSLQPIRYSRASLYGPNIG 120
DB 150 -----EPNNYRTMHGRAVNGSQLGKDYIQLKSLQPIRYSRASLYGPNIG 195
QY 121 RPKNVIALDGFEMKVGSTVDATVQHCYIDGRVVKVMDFLKTRLLDLSQIRKIQKV 180
DB 196 RPKNVIALDGFEMKVGSTVDATVQHCYIDGRVVKVMDFLKTRLLDLSQIRKIQKV 255
QY 181 VNTYTPGKKIWLGVVTTTSGAGTNNLSDSYAAGFLWNLNTGLMANGQIDVIRHSHFFDHG 240
DB 256 VNTYTPGKKIWLGVVTTTSGAGTNNLSDSYAAGFLWNLNTGLMANGQIDVIRHSHFFDHG 315
QY 241 YNHLVDQNFPLDPYWLSSLYLRLGPKVLAVHVAGLQKRPGRVIRDKLRIYAHCTNH 300
DB 316 YNHLVDQNFPLDPYWLSSLYLRLGPKVLAVHVAGLQKRPGRVIRDKLRIYAHCTNH 375
QY 301 HNNYVGRSTTLFIINLHRSRKKIKLAGTLRDKLVHOYLQPYGQGLKSKSVQLNGQPL 360
DB 376 HNNYVGRSTTLFIINLHRSRKKIKLAGTLRDKLVHOYLQPYGQGLKSKSVQLNGQPL 435
QY 361 VMVDDGTLPKLPRLRAGRTLVIPTMTGFFVYKNNNALACRYR 405
DB 436 VMVDDGTLPKLPRLRAGRTLVIPTMTGFFVYKNNNALACRYR 480
RESULT 4
ID Q9QZF8 PRELIMINARY; PRT; 536 AA.
AC Q9QZF8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HEPARANASE.
GN HEP.
OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Podyma K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;
RT "Heparanase from parathyroid cell line."
RL Submitted (Sep-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF184967; AAF04563.1; -
SQ SEQUENCE 536 AA; 60569 MW; 6208B1FD9EE28421 CRC64;

Query Match 38.18; Score 810.5; DB 11; Length 536;
Best Local Similarity 40.68; Pred. No. 4.6e-61;
Matches 170; Conservative 66; Mismatches 114; Indels 69; Gaps 6;

Qy 42 DVMLELQREKAAQHLVLLKEOF-----SNTYS-----NLIL- 73
Db 128 DVLRKLQNEWFFQ-ELLLRQYQREFKNSYSSVDMLYSFAKCSRLDLIFGLNALLR 186
Qy 74 -----TEPNNRTMHGRAVNGSOLGKDYIQLKSL 103
Db 187 TPDLRWNSNAQLLLNYCSSKGYNISWELGNEPNSFWKKAQISIDGLQLGEDFVELHKL 246

Qy 104 QPIRIYSRSLYGNIGPRKNVITALLDGFPMKVAGSTVDATWQHCYIDGRVVKVMDFLK 163
Db 247 QK-SAFQNAKLYGPDIGQPRGKTWKLLRSFLKAGGEVIDSLTWHRYLNGRVATKREDFLS 305

Qy 164 TRLDLTLSQIRKOKVNTVTPGKKIWLEGVTTSSAGTNNLSDSYAAGFLWLTGML 223
Db 306 SDVLDTFILSVQKILKVTKEMTGPKKWLGETSSAYGGAPLLSTFAAGFMWLDKLGLS 365

Qy 224 ANQIDIVVIRHSFFDHGYNHLVDQNFPLPDYWLSSLKRLIGPKVLAHVAGLQKRP 283
Db 366 AOLGTEVVMQVFEGAGNYHLVDENFELPDYWLSSLKRLIGPKVLAHVAGLQKRP 420

Qy 284 GRVTRDKLRIYAHCTNHNHNVVRSITFLINLHRSRKKKLAGTLDKDLVHQLLPY 343
Db 421 ----RSKLRYLVHCTNHNHNVVRSITFLINLHRSRKKKLAGTLDKDLVHQLLPY 476

Qy 344 GQGLKSKSVOLNGOPLVMDVDTLPKLPRLPRAGRTLVIPPPVMTGFFVKNVNALAC 402
Db 477 GSDGLSKSVOLNGOPLVMDVDTLPKLPRLPRAGRTLVIPPPVMTGFFVKNVNALAC 535

RESULT 5
QY251 ID QY251 PRELIMINARY; PRT; 543 AA.
AC QY251;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)
DE HEPARANASE.
GN HPA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA MEDLINE=99321249; PubMed=10395326;
RA Hullett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,
RA Parish C.R.;
RT "Cloning of mammalian heparanase, an important enzyme in tumor
RT invasion and metastasis."
RL Nat. Med. 5:803-809(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Vlodavsky I., Friedman Y., Elkin M., Aingorn H., Atzmon R.,
RA Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I.,
RA Spector L., Pecker I.;
RT "Mammalian heparanase: a novel gene involved in tumor progression and

RT metastasis.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
EX MEDLINE=99377052; PubMed=10446189;
RA Toyoshima M., Nakajima M.;
RT "Human heparanase. Purification, characterization, cloning, and
RT expression."
RN J. Biol. Chem. 274:24153-24160(1999).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=99333379; PubMed=10405343;
RA Kussie P.H., Hulmes J.D., Ludwig D., Patel S., Navarro E.C.,
RA Seddon A.P., Giorgio N.A., Bohlen P.;
RT "Cloning and Functional Expression of a Human Heparanase Gene."
RL Biochem. Biophys. Res. Commun. 261:183-187(1999).
DR EMBL; AF165154; AAD45379.1; -
DR EMBL; AF144325; AAD41342.1; -
DR EMBL; AF155510; AAD45491.1; -
DR EMBL; AF152376; AAD45669.1; -
SQ SEQUENCE 543 AA; 61176 MW; AD262EC267334AB2 CRC64;

Query Match 37.4%; Score 795.5; DB 4; Length 543;
Best Local Similarity 37.8%; Pred. No. 9e-60;
Matches 179; Conservative 66; Mismatches 139; Indels 89; Gaps 10;

Qy 1 NPAKSR-GGPGDYVL-----KNYEDDIYRSOVALDKQCKTAQ-HPDVMLEL 47
Db 88 SPAYLRFGGTKTDFLIFDPKKESTFEERSYQSQVNQDI-----CKYGSIPDPVEEKL 140

Qy 48 OREKAAQHLVLLKEOF-----SNTYS-----NLIL----- 73
Db 141 RLEWYQEQOL-LREHYQKFKNSYSSSDVLYTFANCSDLDLIFGLNALRTADLOW 199

Qy 74 -----TEPNNRTMHGRAVNGSOLGKDYIQLKSLLPY 109
Db 200 NSSNAQLLLDYCSSKGYNISWELGNEPNSFLKADIFINGSOLGDEFIQLHKLARK-STF 258

Qy 110 SRASLYGNIGPRKNVITALLDGFPMKVAGSTVDATWQHCYIDGRVVKVMDFLKTLDT 169
Db 259 KNAKLYGPDVGPQPRKTAKMLKSLKAGGEVIDSVTWHRYLNGRTATREDFLNDVLDI 318

Qy 170 LSDQIRKIQKVVNTYTPGKKIWLEGVTTSSAGTNNLSDSYAAGFLWLTGMLANQGIT 229
Db 319 FTSSVQKVFQVVESTRPKKKWLGETSSAYGGAPLLSDTFAAGFMWLDKLGSLARMGLE 378

Qy 230 VVIRHSFFDHGYNHLVDQNFPLPDYWLSSLKRLIGPKVLAHVAGLQKRPGRVIRD 289
Db 379 VVMQVFEGAGNYHLVDENFELPDYWLSSLKRLIGPKVLAHVAGLQKRPGRVIRD 429

Qy 290 KLRIYAHCTNHNHNVVRSITFLINLHRSRKKKLAGTLDKDLVHQLLPYGOEGLK 349
Db 430 KURVYLHCTNTDNPYKEGDLTLYALNHNVTYKRLPYPPFSNKKQVDRKLLRPLGPHGLL 489

Qy 350 SKSVOLNGOPLVMDVDTLPKLPRLPRAGRTLVIPPPVMTGFFVKNVNALAC 402
Db 490 SKSVOLNGOPLVMDVDTLPKLPRLPRAGRTLVIPPPVMTGFFVKNVNALAC 542

RESULT 6
QYUL39 ID QYUL39 PRELIMINARY; PRT; 545 AA.
AC QYUL39;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE HEPARANASE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

QY 195 VTTSGAGTNLSDSYAAGFLWNTLGLMLANOGIDVIRHSEFFDHGYNHLVDQNFPLPD 254
Db 324 TGSAYGGAPQLSNTYVAGFMWLDKLAARRGIDVVMRQVSFGAGSYHLVDAGFKPLPD 383
QY 255 YWLSLLYKRLGPKVLAVHVAGLQKPPRGVIRDKLRIYAHCTNHHNHNHVRGSGITLFI 314
Db 384 YWLSLLYKRLVTRVLQNSVEQADAR-RP-----RVYLCINPRHPKRVGSDVTILFA 434
QY 315 INLHRSRKIKLAGTLRDLKLVHQLYQYQGLSKSVQNLGQPLVMVDDGTLPKLKPR 374
Db 435 LNLNVTSQSLQPKOLWKSVDVYLLPHGKDSILSREVQLNGRLQLQVDDTLPALHEM 494
QY 375 PLRAGRTLVIPVPMVFFVKNVNALAC 402
Db 495 ALAPGSLGLPAFSYGFYVIRNAKAIAC 522

RESULT 9
Q9SDA1 PRELIMINARY; PRT; 521 AA.

AC Q9SDA1;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HYPOTHEICAL 57.8 KDA PROTEIN.
GN F13G24.30.

OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Van Der Schueren J., Chuang Y.J., Voet M., Robben J.,
RA Volckaert G., Bancroft I., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL133421; CAB62595.1; -.
DR InterPro; IPR001254; Trypsin.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 521 AA; 57831 MW; 07D8664A4B305CC2 CRC64;

Query Match 14.1%; Score 301; DB 10; Length 521;
Best Local Similarity 28.2%; Pred. No. 2.6e-17;
Matches 105; Conservative 50; Mismatches 127; Indels 90; Gaps 15;

QY 84 GRAVNGSQLGRDYIQLKSLLOPIRIYSRASLYGPNIGRPKNVIALLDGF-----M 134
Db 186 GASVSAELYGKDLVLKDVIN--KVYKNSWLHKPLVAP-----GGFYEQQWTKLL 235
QY 135 KVAG-STVDVATVWHQCYIDGR-----VVKVMDFLKTLRLDLSQIRKIQKVVNTYTPG 187
Db 236 EISGSPVVDVVT-HHIYNLGSGNDPALVKKIMD---PSYLSQVSKTFKDVNOTIQEHGP- 290
QY 188 KKIWLEGVVTTSAG-----GTNNLSDSYAAGFLWNTLGLMLANOGIDVIRHSEFFDHGYNH 243
Db 291 ---WASPWVGSGGAYNSGGRHVSOTFIDSFYLDQLGMSARHNTKVYCRQTLVGGFYGL 347
QY 244 LVDQNFNPLDPYLSLLYKRLGPKVLAVHVAGLQKPPRGVIRDKLRIYAHCTNHHNH 303
Db 348 LEKGTFFVNPDPYSALLHRLMGKGLAVQTDG-----PPQLRVYAHCSKG--- 393
QY 304 NVYRGSTILFIINLH-----RSRKKIKLAGTLRDLKLV----- 335
Db 394 ---RAGVTLLNLINSQSDFTVSVSNGINVLNAESRKKKSLDTLKRPFSWIGSKASDG 450
QY 336 ---HOYLLQPYQGEQ-LKSKSVQLNGQPLVMVDDGTLPKLKPRLRAGRTLVIPVPMG 390
Db 390 ---HOYLLQPYQGEQ-LKSKSVQLNGQPLVMVDDGTLPKLKPRLRAGRTLVIPVPMG 390

Db 451 YLNREYHLTP--ENGVLRSKTMVLNGSKLSKPTATGDIPSLPVRKSVNSPLNVLPLSMS 508
QY 391 FFVVKVNNALAC 402
Db 509 FIVLPNFDASAC 520

RESULT 10
Q9FF10 PRELIMINARY; PRT; 543 AA.

AC Q9FF10;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE SIMILARITY TO HEPARANASE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=COLUMBIA;
RX MEDLINE=97471969; PubMed=9330910;
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
RT features of the 1.6 Mb regions covered by twenty physically assigned
RT pl clones";
RL DNA Res. 4:215-230(1997).
DR EMBL; AB005249; BAB09947.1; -.
DR InterPro; IPR001254; Trypsin.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
SQ SEQUENCE 543 AA; 60250 MW; 0FA2248948282FF6 CRC64;

Query Match 14.1%; Score 301; DB 10; Length 543;
Best Local Similarity 28.2%; Pred. No. 2.8e-17;
Matches 105; Conservative 50; Mismatches 127; Indels 90; Gaps 15;

QY 84 GRAVNGSQLGRDYIQLKSLLOPIRIYSRASLYGPNIGRPKNVIALLDGF-----M 134
Db 208 GASVSAELYGKDLVLKDVIN--KVYKNSWLHKPLVAP-----GGFYEQQWTKLL 257
QY 135 KVAG-STVDVATVWHQCYIDGR-----VVKVMDFLKTLRLDLSQIRKIQKVVNTYTPG 187
Db 258 EISGSPVVDVVT-HHIYNLGSGNDPALVKKIMD---PSYLSQVSKTFKDVNOTIQEHGP- 312
QY 188 KKIWLEGVVTTSAG-----GTNNLSDSYAAGFLWNTLGLMLANOGIDVIRHSEFFDHGYNH 243
Db 313 ---WASPWVGSGGAYNSGGRHVSOTFIDSFYLDQLGMSARHNTKVYCRQTLVGGFYGL 369
QY 244 LVDQNFNPLDPYLSLLYKRLGPKVLAVHVAGLQKPPRGVIRDKLRIYAHCTNHHNH 303
Db 370 LEKGTFFVNPDPYSALLHRLMGKGLAVQTDG-----PPQLRVYAHCSKG--- 415
QY 304 NVYRGSTILFIINLH-----RSRKKIKLAGTLRDLKLV----- 335
Db 416 ---RAGVTLLNLINSQSDFTVSVSNGINVLNAESRKKKSLDTLKRPFSWIGSKASDG 472
QY 336 ---HOYLLQPYQGEQ-LKSKSVQLNGQPLVMVDDGTLPKLKPRLRAGRTLVIPVPMG 390
Db 473 YLNREYHLTP--ENGVLRSKTMVLNGSKLSKPTATGDIPSLPVRKSVNSPLNVLPLSMS 530

QY 391 FFVVKVNNALAC 402
Db 531 FIVLPNFDASAC 542

RESULT 11
Q9FLK8 PRELIMINARY; PRT; 516 AA.

ID Q9FLK8
AC Q9FLK8;

DT	01-MAR-2001 (TrEMBLrel. 16, Created)	RC	STRAIN=COLUMBIA;
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)	RA	Kaneko T., Kato T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)	RA	Tabata S.;
DE	SIMILARITY TO HEPARANASE.	RT	"Structural analysis of Arabidopsis thaliana chromosome 5. XI. ";
OS	Arabidopsis thaliana (Mouse-ear cress).	RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	DR	EMBL; AB028613; BAB10787.1; -
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;	SQ	SEQUENCE 536 AA; 59654 MW; 24202B8E82F3DB0E CRC64;
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.		
OX	NCBI_TaxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=COLUMBIA;		
RA	MEDLINE=98290546; PubMed=9628582;		
RA	Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,		
RA	Tabata S.;		
RT	"Structural analysis of Arabidopsis thaliana chromosome 5. IV.		
RT	Sequence features of the regions of 1,456,315 bp covered by nineteen		
RT	physically assigned P1 and TAC clones."		
RL	DNA Res. 5:41-54(1998).		
RL	EMBL; AB010073; BAB08480.1; -		
DR	InterPro: IPR001917; AminoTransf_2.		
DR	PROSITE: PS00599; AA_TRANSF_CLASS_2; UNKNOWN_1.		
SQ	SEQUENCE 516 AA; 57414 MW; 6943E6D3D89FB93C CRC64;		
Query Match	12.1%; Score 258.5; DB 10; Length 516;		
Best Local Similarity	23.7%; Pred. No. 1.2e-13;		
Matches	94; Conservative 63; Mismatches 130; Indels 109; Gaps 14;		
QY	78 NYRTMHGRAVN---GSQL-----GKDYIQLKSLLOPIRIYSRASLYGPNI 119		
Db	119 NYTVSKGYALDSWFEFNELSGSIWASVSVELYGRDLIVLKNVKNYKNSRT----- 210		
QY	120 GRPRKNVIALLDGF-----MKVAGSTVDVAVTWQHICYIDG-----RVVKVMDFLKT 164		
Db	211 ----KPLVWAPGFFEEQWYSELLRLSGPGVDLVLTHTIYNLGNPDNPKLVNKILD---P 263		
QY	165 RLIDTLDQIRKIQKVVNNTYTPGKKIWLGVVTVTSAGFTNN-----LSDSYAAGFLWLNT 219		
Db	264 NYLSGISELFANVQTOIEQHPMAAAW-----GEAGGAFNSGGRGVSEFFINSFWYLDQ 318		
QY	220 LGLMANGQIDVIRHSFFDHGYNHLVDQNFNPLDPYWLKRLGPKVLAVHVGAGLQR 279		
Db	319 LGTSSRHNTKVCYQALVGSFGYGLLEKETFPVNPDPYISALLNHLRMKGILGVQTTA--- 375		
QY	280 KPRGVRIRDKLRIYAHCTHHNNHNVYRGSITLFIINLHR-----S 320		
Db	376 -----SEYLAYVHCSCR-----RAGITILLINLSKHTTFTTAVSNGVKVVLQAES 421		
QY	321 RKKIKLAGTLRDLKLV-----HOYLLQPYCQEG-LKSKSVQLNGOPLVMVDVG 366		
Db	422 MKRKSFLETIKSVSWGNKASDGYLNREYHLSP--KDGDLRSKIMLLNGRPLVPTATG 479		
QY	367 TLPELKPRLPRAGRTLVIPVPTMGFFVVKVNNALAC 402		
Db	480 DIPKLEPVRHGKSPVYINPLSIFVLPTFDAPAC 515		
RESULT 12			
Q9FZP1	PRELIMINARY; PRT; 536 AA.		
AC	Q9FZP1;		
DT	01-MAR-2001 (TrEMBLrel. 16, Created)		
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	SIMILARITY TO HEPARANASE.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.		
OX	NCBI_TaxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
Query Match	10.1%; Score 215.5; DB 10; Length 527;		
Best Local Similarity	28.5%; Pred. No. 5.9e-10;		
Matches	70; Conservative 38; Mismatches 99; Indels 39; Gaps 10;		
QY	64 FSNTYSNLLITERNYR-----TMHGRAVNSQLGKDYIQLKSLLOPIRIYSR 111		
Db	111 YTNAESFIRTAENNYTIDGWELGNELCGSGVGARVGANQYAITNLNRIVN--RVYKN 234		
QY	112 AS-----LYGNIGRPRKNVIALLDGFMKV-----AGSTVDVAVTWQHICY-----IDG 153		
Db	235 VSPMPLVIGPG-----GFFEVDWFTTEYLKKAENSLNATT--RHIYDLGPGVDE 280		
QY	154 RVVKVMDFLKRLDITLSDQIRKIQKVVNNTYTPGKKIWLGVVTVTSAG----GTNNLSDS 209		
Db	281 HLIE--KILNPSYLDQEAQFSRSLKNIKKNSSTKAVAW----VGESGGAYNNGRNLYSNA 334		
QY	210 YAAQFLMINTLGLMANGQIDVIRHSFFDHGYNHLVDQNFNPLDPYWLKRLGPKV 269		
Db	335 FVYSEFWLDQLGMASLYDTKTYCROSLIGGNYGLLNTTFTPNPDYISALIWRLMGRA 394		
QY	270 LAVHVGAGLQRKPRGVRIRDKLRIYAHCTHHNNHNVYRGSITLFIINLHRSR---KKIKL 326		
Db	395 LFTTFSGTK-----KIRSYTHCARQSK-----GITVLMNLDNTTTVVAKVEL 437		
QY	327 AGTLRDLKLVHOYLLQPY-----GOEG-----LKSKSVQLNGOPLVM 362		
Db	438 NNSP--SLRHTKHKMSYKASSQLFEGGPNGVIOREYHLTAQDGNLHLSQTMLLNGNALQV 495		
QY	363 VDDGTLPKLPRLPRAGRTLVIPVPTMGFFVVKVNNALAC 402		
Db	496 NSMGDLPPDIEPIHINSTEPIITAIYSIVFVHMRNVVVPAC 535		
RESULT 13			
Q9LRC8	PRELIMINARY; PRT; 527 AA.		
ID	Q9LRC8;		
AC	Q9LRC8;		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	BETA-GLUCURONIDASE.		
GN	SGUS.		
OS	Scutellaria baicalensis.		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
OC	Asteridae; euasterids I; Lamiales; Lamiaceae; Scutellaria.		
OX	NCBI_TaxID=65409;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=20418130; PubMed=10858442;		
RA	Sasaki K., Taura F., Shoyama Y., Morimoto S.;		
RT	"Molecular Characterization of a Novel beta-Glucuronidase from		
RT	Scutellaria baicalensis Georgi.;"		
RL	J. Biol. Chem. 275:27466-27472(2000).		
DR	EMBL; AB040072; BAA97804.1; -		
DR	InterPro: IPR001179; FKBP_PPase.		
DR	PROSITE: PS00453; FKBP_PPase_1; UNKNOWN_1.		
SQ	SEQUENCE 527 AA; 58772 MW; A5DE7C423F2A1E2B CRC64;		


```
QY 112 ASLYGNIGRPRKNVIALLDGFMKVAGSTVDATWQHCVID-----GRVVKVMDFL 162
Db 145 WRLMGADIS-----AGSSADET---KRYVMSKDLNTAFGWTQFANMLP 185
QY 163 KTRLLDTLSDQIRKIQKVVVNTYTPGKKIWLEGVVTTSAGGTNNLSDSYAAGFLWLNTLGM 222
Db 186 KSSLSGLYSDPALRTLQQQRPV---LWL--TLPEERSQRLVGDETTDALRWAQTMGD 240
QY 223 LANQGDVIVIRHSFFDHGYNHLVD---QNFNPLPDYWLSSLYKRLIGPKVLAVHVAGLQR 279
Db 241 AAASGFDVIFKRM-----NLVDFERPNS----LYVTALEFKVMGSRVFFA----- 282
QY 280 KPRPGRVIRDKLRIVAHCTNHHNNHNVYRGSITLFIINLHRSRKKI--KLAGTLRDKLVHQ 337
Db 283 --RPLNAPSPNKLYTHCA-----NAVSGGLAFMVVNTTEEPTTIVKSTSLSSSEIWQ 335
QY 338 YLQPYGQGLKSKSVQLNGQPLVMYDDGTL-PELKP-----RPLRAGRPLVIPPTMGFF 392
Db 336 YVLTGHDQR-----VOLNNVRLHL--NTTLRPLIKPIDPTKPLQ----LITPSMAVSEFW 383
QY 393 VYKNVNALACRY 404
Db 384 VLPDVNLEHCQF 395
```

Search completed: July 30, 2002, 08:30:49
Job time: 1001 sec


```
Db 88 SPAYLRFGGTKTDFLIFDPKKKESTFEERSYMQSQVNQDI-----CRYGSIPPDVEEKL 140
Qy 48 OREKAAQMHLLVLLKKEOF-----SNTYS-----NLIL----- 73
Db 141 RLEWYPQEQOL-LLREHYQKKFNSTYSRSSVDVLYTFANCSDLDFLFGNALLRTADLOW 199
Qy 74 -----TEPNNYRTMHGRAVNGSOLGKDYIQLKSLQPIRIY 109
Db 200 NSSNAOALLDYCSSKGYNISWELGNEPNSFLKKAIDIFINGSOLGEDYIQLHKLRLK-STF 258
Qy 110 SRASLYGPNIGPRKNVIALDGFEMKAVAGSTVDATWQHCHYIDGRVVKVMDFLKTRLLDT 169
Db 259 KNAKLYGPDVGPQPRRTAKMLKSLFKAGGEVDSVTWHHYLLNGRTATREDFLNPDVLDI 318
Qy 170 LSDQIRKIQKVVNTYTPGKKIWLKQF-----SNTYS-----NLIL----- 73
Db 319 FISSVQKVFQVVESTPRGKKVWLGSETSSAYGGAPLLSDTFAAGFMWLDKGLSARMGIE 378
Qy 230 VVIRHSFFDHGYNHLVDQNFNPLDPYWLSSLYKRLIGPKVLAVHVAGLQKRPGRVIRD 289
Db 379 VVMQVFFGAGNYHLVDENFDPLDPYWLSSLFKLVGTVKVLMSVQSGSKRR----- 429
Qy 290 KLRIYAHCTNHNHNHNVYRGSITLFIINLHRSRKKIKLAGTLRDKLVHQLVLOPYGOEGLK 349
Db 430 KURVYLHCTNTDNPRYKEGDLTLYAINLHNVTYKYLRLPYFNSKNQVDKYLLRPLGPHGLL 489
Qy 350 SKSVOLNGQLVMDVDDGTLPELKPRLPRAGRTLVIPTVMTGFFVVKVNNALAC 402
Db 490 SKSVOLNGLTLKMWDDQTLPLLMEXPLRPGSSGLGPAFSYSEFVIRNAKVAAC 542

RESULT 4
US-09-260-038B-2
; Sequence 2, Application US/09260038B
; Patent No. 6348344
; GENERAL INFORMATION:
; APPLICANT: Maty Ayal-Hershkovitz et al.
; TITLE OF INVENTION: GENETICALLY MODIFIED CELLS AND METHODS FOR
; EXPRESSING RECOMBINANT HEPARANASE
; AND METHODS OF PURIFYING SAME
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
; STREET: 2001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted to
; an ASCII file
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/260,038B
; FILING DATE: 02-Mar-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/071,618
; FILING DATE: May 1, 1998
; APPLICATION NUMBER: 09/071,739
; FILING DATE: May 1, 1998
; APPLICATION NUMBER: 08/922,180
; FILING DATE: September 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedman, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 910/16
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 972-3-5625553
```

```
; TELEFAX: 972-3-5625554
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-260-038B-2

Query Match 37.6%; Score 799.5; DB 4; Length 543;
Best Local Similarity 38.1%; Pred. No. 6e-78;
Matches 180; Conservative 65; Mismatches 139; Indels 89; Gaps 10;

Qy 1 NPAKSR-GGPGPDYLL-----KNYEDDIYRSVDVALDKQKCKTAQ-HPDVMLEL 47
Db 88 SPAYLRFGGTKTDFLIFDPKKKESTFEERSYMQSQVNQDI-----CRYGSIPPDVEEKL 140
Qy 48 OREKAAQMHLLVLLKKEOF-----SNTYS-----NLIL----- 73
Db 141 RLEWYPQEQOL-LLREHYQKKFNSTYSRSSVDVLYTFANCSDLDFLFGNALLRTADLOW 199
Qy 74 -----TEPNNYRTMHGRAVNGSOLGKDYIQLKSLQPIRIY 109
Db 200 NSSNAOALLDYCSSKGYNISWELGNEPNSFLKKAIDIFINGSOLGEDYIQLHKLRLK-STF 258
Qy 110 SRASLYGPNIGPRKNVIALDGFEMKAVAGSTVDATWQHCHYIDGRVVKVMDFLKTRLLDT 169
Db 259 KNAKLYGPDVGPQPRRTAKMLKSLFKAGGEVDSVTWHHYLLNGRTATREDFLNPDVLDI 318
Qy 170 LSDQIRKIQKVVNTYTPGKKIWLKQF-----SNTYS-----NLIL----- 73
Db 319 FISSVQKVFQVVESTPRGKKVWLGSETSSAYGGAPLLSDTFAAGFMWLDKGLSARMGIE 378
Qy 230 VVIRHSFFDHGYNHLVDQNFNPLDPYWLSSLYKRLIGPKVLAVHVAGLQKRPGRVIRD 289
Db 379 VVMQVFFGAGNYHLVDENFDPLDPYWLSSLFKLVGTVKVLMSVQSGSKRR----- 429
Qy 290 KLRIYAHCTNHNHNHNVYRGSITLFIINLHRSRKKIKLAGTLRDKLVHQLVLOPYGOEGLK 349
Db 430 KURVYLHCTNTDNPRYKEGDLTLYAINLHNVTYKYLRLPYFNSKNQVDKYLLRPLGPHGLL 489
Qy 350 SKSVOLNGQLVMDVDDGTLPELKPRLPRAGRTLVIPTVMTGFFVVKVNNALAC 402
Db 490 SKSVOLNGLTLKMWDDQTLPLLMEXPLRPGSSGLGPAFSYSEFVIRNAKVAAC 542

RESULT 5
US-09-181-336-13
; Sequence 13, Application US/09181336A
; Patent No. 6242238
; GENERAL INFORMATION:
; APPLICANT: FREEMAN, Craig Geoffrey
; APPLICANT: PARISH, Richard
; APPLICANT: HAMDORF, Brenton James
; APPLICANT: HULETT, Mark Darren
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN
; TITLE OF INVENTION: ENDOGLUCURONIDASE AND USES THEREFOR
; FILE REFERENCE: 032505-004
; CURRENT APPLICATION NUMBER: US/09/181,336A
; CURRENT FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: AU PP0062
; EARLIER FILING DATE: 1997-10-28
; EARLIER APPLICATION NUMBER: AU PP0812
; EARLIER FILING DATE: 1997-12-09
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Homo sapiens
```

US-09-181-336-13									
Query Match 37.4%; Score 795.5; DB 4; Length 543;									
Best Local Similarity 37.8%; Pred. No. 1.6e-77;									
Matches 179; Conservative 66; Mismatches 139; Indels 89; Gaps 10;									
QY	1	NPAKSR-GGPGDYVL-----KNYEDDIVRSVDVALDKGCKIAQ-HPDVMLEL	47						
DB	88	SPAYLRFGGTKDFLIFDPKXESTFEERSYQSQVNQDI-----CKYGSIPPDVEEKL	140						
QY	48	QREKAAQMHVLVLLKEQF-----SNTYS-----NLIL-----	73						
DB	141	RLWPYQEQ-LLRHYQKFKFNSTYSSVDVLYTFANCGLDLIFGLNALLRTADLOW	199						
QY	74	-----TEPNYRTMHGRAVNSQLGKDYIQLKSLQPIRY	109						
DB	200	NSSNAQLLLDYCSSKGYNTSWELGNEPNSFLKKADIFINGSQLGEDFIQHLKLRK-STF	258						
QY	110	SRLSYGPNIGRPRKNVIALLDGFMKRVAGSTVDATVWQHCYIDGRVVKVMDFLKTRLLDT	169						
DB	259	KNAKLYGPDVGQPRRTAKMLKSLFKAGGEVDSVTWHYVYLNRTATREDPLNPDVLDI	318						
QY	170	LSQDIKRIKQVNTYTPGKKIWLEGVVTT'SAGTNNLSDSYAAGFLWLTGLMLANQGD	229						
DB	319	FISVQKVFQVVESTRPGKKVWLGETSSAYGGAPLLSDTFAAGFMWLDKGLSARMGIE	378						
US-09-181-336-17									
Query Match 36.8%; Score 779; DB 4; Length 380;									
Best Local Similarity 46.3%; Pred. No. 5.6e-76;									
Matches 152; Conservative 56; Mismatches 110; Indels 10; Gaps 2;									
QY	75	EPNNYRTMHGRAVNSQLGKDYIQLKSLQPIRYISRLSYGPNIGRPRKNVIALLDGFM	134						
US-09-181-336-17									
Query Match 35.3%; Score 752.5; DB 4; Length 532;									
Best Local Similarity 38.0%; Pred. No. 7.2e-73;									
Matches 172; Conservative 62; Mismatches 130; Indels 89; Gaps 10;									
QY	1	NPAKSR-GGPGDYVL-----KNYEDDIVRSVDVALDKGCKIAQ-HPDVMLEL	47						
DB	88	SPAYLRFGGTKDFLIFDPKXESTFEERSYQSQVNQDI-----CKYGSIPPDVEEKL	140						
QY	48	QREKAAQMHVLVLLKEQF-----SNTYS-----NLIL-----	73						
DB	141	RLWPYQEQ-LLRHYQKFKFNSTYSSVDVLYTFANCGLDLIFGLNALLRTADLOW	199						
QY	74	-----TEPNYRTMHGRAVNSQLGKDYIQLKSLQPIRY	109						
DB	200	NSSNAQLLLDYCSSKGYNTSWELGNEPNSFLKKADIFINGSQLGEDFIQHLKLRK-STF	258						
QY	110	SRLSYGPNIGRPRKNVIALLDGFMKRVAGSTVDATVWQHCYIDGRVVKVMDFLKTRLLDT	169						
DB	259	KNAKLYGPDVGQPRRTAKMLKSLFKAGGEVDSVTWHYVYLNRTATREDPLNPDVLDI	318						
QY	170	LSQDIKRIKQVNTYTPGKKIWLEGVVTT'SAGTNNLSDSYAAGFLWLTGLMLANQGD	229						
DB	319	FISVQKVFQVVESTRPGKKVWLGETSSAYGGAPLLSDTFAAGFMWLDKGLSARMGIE	378						
US-09-181-336-15									
Query Match 35.3%; Score 752.5; DB 4; Length 532;									
Best Local Similarity 38.0%; Pred. No. 7.2e-73;									
Matches 172; Conservative 62; Mismatches 130; Indels 89; Gaps 10;									
QY	1	NPAKSR-GGPGDYVL-----KNYEDDIVRSVDVALDKGCKIAQ-HPDVMLEL	47						
DB	88	SPAYLRFGGTKDFLIFDPKXESTFEERSYQSQVNQDI-----CKYGSIPPDVEEKL	140						
QY	48	QREKAAQMHVLVLLKEQF-----SNTYS-----NLIL-----	73						
DB	141	RLWPYQEQ-LLRHYQKFKFNSTYSSVDVLYTFANCGLDLIFGLNALLRTADLOW	199						
QY	74	-----TEPNYRTMHGRAVNSQLGKDYIQLKSLQPIRY	109						
DB	200	NSSNAQLLLDYCSSKGYNTSWELGNEPNSFLKKADIFINGSQLGEDFIQHLKLRK-STF	258						
QY	110	SRLSYGPNIGRPRKNVIALLDGFMKRVAGSTVDATVWQHCYIDGRVVKVMDFLKTRLLDT	169						
DB	259	KNAKLYGPDVGQPRRTAKMLKSLFKAGGEVDSVTWHYVYLNRTATREDPLNPDVLDI	318						
QY	170	LSQDIKRIKQVNTYTPGKKIWLEGVVTT'SAGTNNLSDSYAAGFLWLTGLMLANQGD	229						
DB	319	FISVQKVFQVVESTRPGKKVWLGETSSAYGGAPLLSDTFAAGFMWLDKGLSARMGIE	378						

Db	90	KGG	-----FAKCYE-----	MDLTNNKYAAKI	PHSRVAKPHQREK	IKETIELHRIHHK	140
QY	56	HLVLLKEQFSNTY	NNLILPEPNY	RTMHGRAVNGSOL	GKDYIQLKSLLOP	IRIYSRASL	114
Db	141	HVQYFHYFEDKENIY	ILLEYCSRSM-----	AHILK-----	ARKVLTEPEVRYLQIV	190	
QY	115	YGNIGRPNKNTALIDG	FMKVAGSVTDVATW	QHCYIDGRVVVKVMD	FKTRLLDLSQI	174	
Db	191	SGLYLHQE-----	ILHRDLKGNFPIN-----	EAMELKVGDFGLAA	LEPLEHRR	237	
QY	175	RKQKVVNYP-----	GKKTWLGVV-----	TTSAG-----	GTNNLSDSY-----	AAGF	214
Db	238	RTICGTPNYLSU	PEVLNKQHGCESD	LWALGCVMYMLLGR	PFETNLKETYRCI	REARY	297
QY	215	LWLNTL-----	GMLAQ-----	GIDVTRHSPFDHGY	-----NHILVQNFNPL	DDYWLUS	258
Db	298	TMPSSLLAPAKHLI	ASMLSKNPNDRPSL	DDIIRHDFLOQGF	TPDRLSSCCCHTV	PDFHLS	357
QY	259	LLYKRLIGPKVLAV	HVAVAGLQKRP-----	PCRVTRDKLRTIYA	HTNHHNNHNVGR	SIT	311
Db	359	SDAKNEE-KYAAAI	ECCKQKQAVYD	TNNVSKENEDIV-----	KLRHDLKKTST	408	

```

RESULT 10
US-09-136-282-2
; Sequence 2, Application US/09136282
; Patent No. 6063609
; GENERAL INFORMATION:
; APPLICANT: ANDERSON, KAREN
; APPLICANT: JACKSON, JEFFREY
; APPLICANT: HANSBURY, MICHAEL
; APPLICANT: NERURKAR, SANDHYA
; APPLICANT: ROSHAK, AMY
; APPLICANT: BOUZYK, MARK
; TITLE OF INVENTION: HUMAN SERUM INDUCIBLE KINASE (SNK)
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,282
; FILING DATE: 20-AUG-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/056,112
; FILING DATE: 20-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Prestia, Paul F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH-70231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0700
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 685 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-136-282-2

```

```

Query Match          4.3%; Score 92.5; DB 3; Length 685;
Best Local Similarity 23.2%; Pred. No. 0.43;
Matches      83; Conservative 46; Mismatches 139; Indels    89; Gaps
QY   6 RGGGPDYLYLNVEDDIVRSVDALDKQCKCKTAQHDPDVMLELQREKAAQ-----M 55
DB   90 KGG-----FAKCYE----MTDLTNKKVYAAKIPIHSRVAKPQREKDKEIEHLHILHHK 140
QY   56 HLVLLEKQESNTYSNLILTPEPNRYRTMHGRAVNGSOLGXDYTQLKASLQP-IRIYSRASL 114
DB   141 HVVFHYHFEDKENIYLLEYCSRSSK-----AHILK---ARKVLTEPEVRYYLRQIV 190
QY   115 YGNIGRPKRKNVIALLDGMFKMVAGSTVDVTWOHCVIDGRVKVMDFLKTRLLDTLSDOI 174
DB   191 SGLKYLHEQE-----ILHRDLKLGNFFIN-----EAMELKVGDFGUAAARLEPLEHRR 237
QY   175 RKTKQVVNTYTP-----GKKTLWLEGV--TTSAG-----GTNNLSDSY-----AAGF 214
DB   238 RTICGTGPNYLSPSEVLNKKGHGCESDIWALGVCMYTWMLGRPFPETTNLKETRCIREARY 297
QY   215 LWLNLT-----GMLANO-----GIDVVIHRSFFDHGY--NHILVDQNENPLPDYWLS 258
DB   298 TMSSELLAPAKHLIASMLSKNPEDRPSLDDIIRHDEFFLOGFTPDRLSSSCCHTVPDFHLS 357
QY   259 LLYKRLGPKVLAVHVAGLQRKPR-----PGRVTRDKLRIYAHCTNHNHNINYRGSGIT 311
DB   358 SPAKNFF-KKAAAALFGKKDKKARIYDTHNRVSKEDEDIY-----KLRHDLKKTIST 408

RESULT 11
US-09-272-796-1
; Sequence 11, Application US/09272796
; Patent No. 6207148
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Nell C.
; APPLICANT: Guegler, Karl G.
; APPLICANT: Lal, Preeti
; APPLICANT: Goli, Surya K.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/272,796
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/878,989
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PE-0321 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:

```

SEQUENCE CHARACTERISTICS:

LENGTH: 685 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: HUVENB01
CLONE: 39043
US-09-272-796-1

Query Match

Best Local Similarity 4.3%; Score 92.5; DB 4; Length 685;

Matches 83; Conservative 46; Mismatches 139; Indels 89; Gaps 18;

Qy 6 RGGPGPDYILKNYEDDIVRSVDALDKQCKIAQHPDVMLELQREKAAQ-----M 55
Db 90 KGG-----FAKCYE---MTDLTNKYAAKIIPHSRVAKPHQREKIDKEIHLHRLHK 140
Qy 56 HLVLKKEQFSNTYSLNLIITPEPNNTYTHGRAVNGSOLGKDYIQLKSLQIP-IRIYSRASL 114
Db 141 HVVOFYHYFEDKENIYILLEYCSRSM-----AHILK---ARKVLTEPEVRYLQIV 190
Qy 115 YGPNIGRPRKNVIALLDGFMKVGASTVDATWQHCYIDGRVVKVMDFLKTRLLDLSQI 174
Db 191 SGLKYLHEQE-----ILHRDLKLGNNFFIN-----EAMELKVGDFGLAARLEPLEHRR 237
Qy 175 RKIQKVVNTYTP-----GKKIWLEGVY--TTSAG-----GTNNLSDSY----AAGF 214
Db 238 RTICGTPNLYSPEVLNKGHCESDIWALGCVMTMLLGRPPFTTNLKYRCIREARY 297
Qy 215 LMLNTL-----GMLANO-----GIDVVRHSFFDHGY--NHLVDQNFNPLDPYWLS 258
Db 298 TMSLLAPAKHLIASLMSKNPDRPSLDDIIRHDFLQGTTPDRLSSCCCHTVDPFHL 357
Qy 259 LLYKRLGPKVLAVHVAGLQKPR-----PGRVIRDKLRIYAHCTNHHNNHNVVRSIT 311
Db 358 SPANKFF-KKAAALFGGKDKKARYIDTHNRVSKEDDIY-----KLRLDLKKTST 408

RESULT 12

US-09-505-744-2
Sequence 2, Application US/09505744
Patent No. 6245544
GENERAL INFORMATION:
APPLICANT: Karen M. Anderson
APPLICANT: Mark M. Bouzyk
APPLICANT: Michael J. Hansbury
APPLICANT: Jeffrey R. Jackson
APPLICANT: Sandhya S. Nerurkar
APPLICANT: Amy K. Roshak
TITLE OF INVENTION: HUMAN SERUM INDUCIBLE KINASE (SNK)
FILE REFERENCE: GH-70231-D1
CURRENT APPLICATION NUMBER: US/09/505,744
CURRENT FILING DATE: 2000-02-16
EARLIER APPLICATION NUMBER: 09/136,282
EARLIER FILING DATE: 1998-08-20
EARLIER APPLICATION NUMBER: 60/056,112
EARLIER FILING DATE: 1997-08-20
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 685
TYPE: PRT
ORGANISM: HOMO SAPIENS
US-09-505-744-2

Query Match

Best Local Similarity 4.3%; Score 92.5; DB 4; Length 685;

Matches 83; Conservative 46; Mismatches 139; Indels 89; Gaps 18;

Qy 6 RGGPGPDYILKNYEDDIVRSVDALDKQCKIAQHPDVMLELQREKAAQ-----M 55

Db 90 KGG-----FAKCYE---MTDLTNKYAAKIIPHSRVAKPHQREKIDKEIHLHRLHK 140
Qy 56 HLVLKKEQFSNTYSLNLIITPEPNNTYTHGRAVNGSOLGKDYIQLKSLQIP-IRIYSRASL 114
Db 141 HVVOFYHYFEDKENIYILLEYCSRSM-----AHILK---ARKVLTEPEVRYLQIV 190
Qy 115 YGPNIGRPRKNVIALLDGFMKVGASTVDATWQHCYIDGRVVKVMDFLKTRLLDLSQI 174
Db 191 SGLKYLHEQE-----ILHRDLKLGNNFFIN-----EAMELKVGDFGLAARLEPLEHRR 237
Qy 175 RKIQKVVNTYTP-----GKKIWLEGVY--TTSAG-----GTNNLSDSY----AAGF 214
Db 238 RTICGTPNLYSPEVLNKGHCESDIWALGCVMTMLLGRPPFTTNLKYRCIREARY 297
Qy 215 LMLNTL-----GMLANO-----GIDVVRHSFFDHGY--NHLVDQNFNPLDPYWLS 258
Db 298 TMSLLAPAKHLIASLMSKNPDRPSLDDIIRHDFLQGTTPDRLSSCCCHTVDPFHL 357
Qy 259 LLYKRLGPKVLAVHVAGLQKPR-----PGRVIRDKLRIYAHCTNHHNNHNVVRSIT 311
Db 358 SPANKFF-KKAAALFGGKDKKARYIDTHNRVSKEDDIY-----KLRLDLKKTST 408

RESULT 13

US-08-435-675B-5
Sequence 5, Application US/08435675B
Patent No. 5710250
GENERAL INFORMATION:
APPLICANT: Ellis, Steven Bradley
APPLICANT: Williams, Mark E.
APPLICANT: Harpold, Michael Miller
APPLICANT: Schwartz, Arnold
APPLICANT: Brenner, Robert
TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/435,675B
FILING DATE: 05-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,083
FILING DATE: 28-SEP-1994
APPLICATION NUMBER: US 07/914,231
FILING DATE: 13-JUL-1992
APPLICATION NUMBER: US 07/603,751
FILING DATE: 08-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-53193
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1106 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

Search completed: July 30, 2002, 08:15:07
Job time: 289 sec

[illegible]



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:14:03 ; Search time 141.47 Seconds
(without alignments)
292.857 Million cell updates/sec

Title: US-09-836-461-2_COPY_162_534

Perfect score: 1956
Sequence: 1 KCCKTAQHPDVMLEQREKA.....PVTMGFFVKNVNALACRYR 373

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1956	100.0	534	22 AAB85216	Heparanase-like pr
2	1949	99.6	492	22 AAB84664	Amino acid sequenc
3	1917	98.0	592	22 AAU07424	Human heparanase-1
4	1917	98.0	592	22 AAB85215	Heparanase-like pr
5	1917	98.0	592	22 AAY97632	Human heparanase,
6	1907	97.5	592	22 AAB81062	Human heparanase-2
7	1743	89.1	439	22 AAU07423	Human heparanase-1
8	1743	89.1	480	22 AAU07418	Novel human extrac
9	1743	89.1	480	22 AAB85217	Heparanase-like pr
10	1743	89.1	480	22 AAY97634	Human heparanase,
11	1740	89.0	538	22 AAY97633	Human heparanase,

12	1112	56.9	214	22 AAM99905	Human excretory re
13	1112	56.9	214	22 AAM43704	Human bladder anti
14	805	41.2	380	20 AAY17085	Rat heparanase enz
15	791.5	40.5	543	20 AAY02345	A human heparanase
16	791.5	40.5	543	21 AAB08849	Amino acid sequenc
17	791.5	40.5	543	21 AAY57590	Human heparanase.
18	791.5	40.5	543	21 AAY52990	Human heparanase p
19	791.5	40.5	543	22 AAY97635	Human heparanase p
20	791.5	40.5	592	20 AAY02346	A human heparanase
21	791.5	40.5	592	21 AAB08850	Amino acid sequenc
22	788.5	40.3	530	20 AAY34173	Human pre-prohepar
23	788.5	40.3	543	22 AAB88361	Human heparanase e
24	787.5	40.3	543	20 AAY17082	Human membrane or
25	787.5	40.3	543	22 AAB86206	Human heparanase i
26	787.5	40.3	588	20 AAY30124	A human protein wi
27	784.5	40.1	535	21 AAB08851	A murine heparanas
28	779	39.8	380	20 AAY17084	Mouse heparanase e
29	778	39.8	156	22 AAG55963	Human heparanase-1
30	744.5	38.1	532	20 AAY17083	Seq ID No: 15 of W
31	485	24.8	262	22 AAM24147	Human EST encoded
32	403	20.6	488	22 AAB31469	Amino acid sequenc
33	388	19.8	488	22 AAB31470	Amino acid sequenc
34	383	19.6	488	22 AAB31472	Amino acid sequenc
35	375	19.2	488	22 AAB31471	Amino acid sequenc
36	278	14.2	137	22 AAG55961	Human heparanase-1
37	278	14.2	159	22 AAG55964	Human heparanase-1
38	247	12.6	256	21 AAG13479	Arabidopsis thalia
39	213	10.9	112	22 AAU07425	Human heparanase-1
40	203	10.4	118	22 AAG55962	Human heparanase-1
41	155	7.9	935	22 ABB89219	Drosophila melanog
42	144	7.4	38	20 AAY34186	Human pre-prohepar
43	127	6.5	24	22 AAB85220	Heparanase-like pr
44	127	6.5	25	22 AAB85221	Heparanase-like pr
45	121	6.2	23	22 AAB85219	Heparanase-like pr

ALIGNMENTS

RESULT 1

AAB85216
ID AAB85216 standard; Protein; 534 AA.

AC AAB85216;

DT 07-SEP-2001 (first entry)

DE Heparanase-like protein Hpa2 splice variant #2.

XX Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;
KW cytosolic; neuroprotective; cerebroprotective; immunosuppressive;
KW antipsoriatic; nootropic; antiinflammatory; antiarthritic; antiasthmatic;
KW antidiabetic; antiarteriosclerotic; vulnery.

XX Homo sapiens.

XX WO200146392-A2.

XX 28-JUN-2001.

XX 21-DEC-2000; 2000WO-GB04963.

XX 22-DEC-1999; 99GB-0030392.

XX 07-APR-2000; 2000GB-0008713.

XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.

XX McKenzie EA, Stamps AC, Terrett JA, Tyson KL;

XX WPI; 2001-418056/44.

XX N-PSDB; AAB22672.

XX Novel homologs of heparanase, present in three splice variants, useful

PT for identifying agents that modulate heparanase, useful in the
PT treatment and/or prophylaxis of abnormal levels of heparanase -
XX
XX
XX Claim 1; Fig 2; 97pp; English.
XX
XX The invention provides a homologue to heparanase which is present in
XX three splice variants. The heparanase homologue polypeptide is useful in
XX the treatment of a human or non-human animal or for use in diagnosis.
XX Vectors comprising the heparanase homologue polynucleotides are useful in
XX the transformation or transfection of a prokaryotic or eukaryotic host.
XX The modulators of the polypeptide are useful in the manufacture of a
XX medicament for the treatment and/or prophylaxis of a condition/disease
XX associated with abnormal levels of the heparanase homologue, including
XX cancer, central nervous system (CNS) and neurodegenerative diseases,
XX cardiovascular diseases such as stenosis following angioplasty and
XX atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus,
XX allografts, inflammatory diseases, arthritis, vascular stenosis,
XX tumour growth and progression, asthma, Alzheimer's disease, diabetic
XX retinopathy, wound healing and inflammation. The polypeptide is also
XX useful in diagnosis and research. The present sequence represents the
XX amino acid sequence of the mid-sized splice variant of the heparanase-
XX like protein Hpa2 of the invention.
XX
XX Sequence 534 AA;

Query Match 100.0%; Score 1956; DB 22; Length 534;
Best Local Similarity 100.0%; Pred. No. 1.9e-199;
Matches 373; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KGCKTAQHPDVMLELQREKAAQMHLVLLKEQFSNTYSNLIITPEPNYRTMHGRAVNSQL 60
Db 162 kgciaqhpdvmllelqrekaqmhlvllkeqfsntysnliitepnnyrtmhgravnsgql 221
Qy 61 GKDYIQLKSLLOPIRIYRASLYGNIGRPNKRVIALDGFPMKVGSTVDVWQHCVYID 120
Db 222 gkdylqkllqpriyraslygnigrpnknvialldgfmkvagstvavtwqhcyid 281
Qy 121 GRVVKVMDFLKTRLDLTDSDQIRKIQKVNTYTPGKKIWLGEVVTTSAGGTNNLSDSYAA 180
Db 282 grvkvmdflktrldtldsdqirkikvntytpgkklwlegvvttsaggttnlstdsya 341
Qy 181 GFLWTLTGLMLANOGIDVYIRHSFDDHGYNHLVDQNFNPLPDYWLISLLYKRLIGPKVLAV 240
Db 342 gflwntlgmlangidvyrhsfddhgynhlvdqnfplpdywlslllykrligpkvlav 401
Qy 241 HVAGLQRPGRVIRDKLRIYAHCTNNHNNHYVRSSTLFTIINLHRSRKKIKLAGTLRD 300
Db 402 hvaglrprgrvirdklriyahctnnhnnhnyvrgstlftiinlhrsrrkkiklagtlrd 461
Qy 301 KLVHGYLLQPYGQGLKSKSVOLNGQPLVMVDDGTLPELKPRLRAGRTLVIPTVTMGFF 360
Db 462 klvhgyllqpygqglksksvqlngqplvmvddgtlpeklprlragrtlviptvtmgff 521
Qy 361 VVKNNALACRYR 373
Db 522 vvknnalacryr 534

RESULT 2
AAB84664
ID AAB84664 standard; Protein; 492 AA.
XX
XX AAB84664;
AC
AC Amino acid sequence of human heparanase-like polypeptide.
DT
DT 05-SEP-2001 (first entry)
XX
XX Human; heparanase-like polypeptide; gene therapy; cancer; angiogenesis;
KW trauma; autoimmune disease; skin disease; cardiovascular disease;
KW nervous system disease; inflammation; arthritis; genitalia;
KW male fertility; erectile dysfunction.

XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FT Misc-difference 407 /note= "unspecified residue encoded by KCA"
FT
XX
XX WO200148161-A2.
XX
XX 05-JUL-2001.
XX
XX 18-DEC-2000; 2000WO-EP12909.
XX
XX 23-DEC-1999; 99EP-0125831.
XX (SCHD) SCHERING AG.
XX
XX Siemeister G, Weiss B;
XX
XX WPI: 2001-418259/44.
XX N-PSDB; AAH28347.
XX
XX Human Heparanase-like polynucleotide encoding polypeptides useful for
XX modulating expression of the polypeptide and for treating cancer,
XX cancer metastasis, aberrant angiogenesis by gene therapy technique -
XX
XX Claim 9; Page 30; 30pp; English.
XX
XX The present sequence represents a human heparanase-like polypeptide.
XX Heparanase-like polynucleotides are useful as a source of probes,
XX primers and antisense molecules, and in gene therapy. Heparanase-like
XX polynucleotides and polypeptides are useful for treating several
XX disorders e.g., cancer, cancer metastasis. The oligonucleotides are
XX also useful as diagnostic markers for the diagnosis of disorder such
XX as cancer, cancer metastasis and aberrant angiogenesis. They may also
XX act as diagnostic markers for diagnosis of disorder such as cancer,
XX cancer metastasis and aberrant angiogenesis. The heparanase polypeptides
XX and polynucleotides are also useful for treating trauma, autoimmune
XX diseases, skin diseases, cardiovascular diseases, nervous system
XX diseases, and inflammation including arthritis. Since the polynucleotide
XX is preferentially expressed in male genitalia, modulation of its
XX expression and/or activity may be used for medical intervention in male
XX genitalia function that is male fertility control, erectile dysfunction.
XX
XX Sequence 492 AA;

Query Match 99.6%; Score 1949; DB 22; Length 492;
Best Local Similarity 99.5%; Pred. No. 9.6e-199;
Matches 371; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 KGCKTAQHPDVMLELQREKAAQMHLVLLKEQFSNTYSNLIITPEPNYRTMHGRAVNSQL 60
Db 120 kgciaqhpdvmllelqrekaqmhlvllkeqfsntysnliitepnnyrtmhgravnsgql 179
Qy 61 GKDYIQLKSLLOPIRIYRASLYGNIGRPNKRVIALDGFPMKVGSTVDVWQHCVYID 120
Db 180 gkdylqkllqpriyraslygnigrpnknvialldgfmkvagstvavtwqhcyid 239
Qy 121 GRVVKVMDFLKTRLDLTDSDQIRKIQKVNTYTPGKKIWLGEVVTTSAGGTNNLSDSYAA 180
Db 240 grvkvmdflktrldtldsdqirkikvntytpgkklwlegvvttsaggttnlstdsya 299
Qy 181 GFLWTLTGLMLANOGIDVYIRHSFDDHGYNHLVDQNFNPLPDYWLISLLYKRLIGPKVLAV 240
Db 300 gflwntlgmlangidvyrhsfddhgynhlvdqnfplpdywlslllykrligpkvlav 359
Qy 241 HVAGLQRPGRVIRDKLRIYAHCTNNHNNHYVRSSTLFTIINLHRSRKKIKLAGTLRD 300
Db 360 hvaglrprgrvirdklriyahctnnhnnhnyvrgstlftiinlhrsrrkkiklagtlrd 419
Qy 301 KLVHGYLLQPYGQGLKSKSVOLNGQPLVMVDDGTLPELKPRLRAGRTLVIPTVTMGFF 360
XX
XX

Db 420 klvhqyllqpyggqgkkskvsqngqplvmvddgtlpeklprprlagrtlvipvptmgfy 479

Qy 361 VVKVNALACRYR 373

Db 480 vkvvnalacryr 492

RESULT 3

AAU07424

XX AAU07424 standard; Protein; 592 AA.

AC AAU07424;

XX 18-DEC-2001 (first entry)

XX Human heparanase-like protein splice variant #1.

XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;

KW antiproliferative; cardiac; vasotropic; cerebroprotective; neuropro-

KW cerebroprotective; antibacterial; virucide; fungicide; ophthalmological;

KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;

KW hyperproliferative disorder; neoplasm; cardiovascular disorder;

KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;

KW nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;

KW wound healing; food additive; heparanase.

XX Homo sapiens.

OS

XX WO200179253-A1.

PN 25-OCT-2001.

XX 11-APR-2001; 2001WO-US11643.

PF 18-APR-2000; 2000US-198123P.

PR (HUMA-) HUMAN GENOME SCI INC.

PA

XX Fiscella M, Shi Y, Ebner R, Ruben SM;

PI WPI; 2001-611720/70.

XX N-PSDB; AAS13848.

DR

XX New nucleic acids encoding extracellular matrix polypeptides, for

FT diagnosing, treating, preventing or ameliorating human disorders and

PT disease, such as, autoimmune, hyperproliferative or cardiovascular

PT disorders -

XX Disclosure; Page 14; 308pp; English.

PS

XX The invention relates to novel isolated polynucleotides (I) encoding

CC extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by

CC (I) are used to prevent, treat or ameliorate a medical condition in e.g.

CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They

CC are also used in diagnosing a pathological condition or susceptibility to

CC a pathological condition. The antibodies to the polypeptides can also be

CC used in alleviating symptoms associated with the disorders and in

CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked

CC immunoassays (ELISA). Disorders which are diagnosed or treated

CC include autoimmune diseases e.g. rheumatoid arthritis,

CC hyperproliferative disorders e.g. neoplasms of the breast or liver,

CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders

CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.

CC Alzheimer's disease, infections caused by bacteria, viruses and fungi

CC and ocular disorders e.g. corneal infection. The polypeptides can also

CC be used to aid wound healing and epithelial cell proliferation, to

CC prevent skin aging due to sunburn, to maintain organs before

CC transplantation, for supporting cell culture of primary tissues, to

CC regenerate tissues and in chemotaxis. The polypeptides can also be used

CC as a food additive or preservative to increase or decrease storage

CC capabilities. The present sequence represents the amino acid sequence

CC of human heparanase-like protein, splice variant #1.

XX

SQ Sequence 592 AA;

Query Match 98.0%; Score 1917; DB 22; Length 592;

Best Local Similarity 86.5%; Pred. No. 3.3e-195;

Matches 373; Conservative 0; Mismatches 0; Indels 58; Gaps 1;

Qy 1 KGCKIAQHPDVMLQREKAAQMHLLVLKEQFSNTYSNLILT-----EPNNYRTMHGRAVNGSQLGK 62

Db 162 kgckiaqhpdvmlqrekaaqmhlvllkeqfsntysnliltarsldklynfadcsghl 221

Qy 43 -----EPNNYRTMHGRAVNGSQLGK 62

Db 222 ifalnairmpnnswnssalsllkysaskkyniswelgnepnnyrtmhgravngsqlgk 281

Qy 63 DYIOLKSELQPIRIVSRASLYGPNIGRPKNVIALLDGFMKVAGSTVDVAVTWOHCYIDGR 122

Db 282 dyiolksllqpirlysrasygpnigrpknvialldgfmkvagstvdavtwqhcylodr 341

Qy 123 VVKVMDFLKTRLLDLSQIRKIQKVNTYTPGKKIWIIEGVVTTTSAGGTNNLSDSYAAGF 182

Db 342 vkvvmdflktrlldtsdqirkiqkvvntytpgkkiwilegvvtttsaggtnnlsdsyaagf 401

Qy 183 LWLNTGLMANOGIDVVRHSFFDHGYNHLVDQNFNPDPYWLSSLYKRLIGPKVLAVHV 242

Db 402 lwlntglmanogidvvirhsffdhgynhlvdqnfnpdpywlslykrligpkvlavhv 461

Qy 243 AGLORPKPRGVRIRDKLRIYAHCTNHHNHNHYRGSTITLFIINLHRSKKIKLAGTLRDKL 302

Db 462 aglorkprprgvrirdklriyahctnhhnhnhyrgstifilnlhrrskkiklagtlrdkl 521

Qy 303 VHQYLLQPYGQGLKSKSVQLNGOPLVMVDGTLPELKPRLRAGRTLVIPTVMTGFFVW 362

Db 522 vhyllpygqeglksvqlngqplvmvddgtlpeklprprlagrtlviptvmtgffvw 581

Qy 363 KVNVALACRYR 373

Db 582 knvnalacryr 592

RESULT 4

AAB85215

ID AAB85215 standard; Protein; 592 AA.

XX

AC AAB85215;

XX

DT 07-SEP-2001 (first entry)

XX

DE Heparanase-like protein Hpa2 splice variant #1.

XX

KW Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;

KW cytostatic; neuroprotective; cerebroprotective; immunosuppressive;

KW antipsoriatic; nootropic; antiinflammatory; antiarthritic; antiasthmatic;

XX antidiabetic; antiarteriosclerotic; vulnerary.

OS Homo sapiens.

FX

FT Key Location/Qualifiers

FT Misc-difference 237

FT /label= unknown

FT /note= "encoded by ANC"

PN WO200146392-A2.

XX

PD 28-JUN-2001.

XX

XX 21-DEC-2000; 2000WO-GB04963.

XX

PR 22-DEC-1999; 99GB-0030392.

PR 07-APR-2000; 2000GB-0008713.

XX

PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.

XX

Qy 303 VHQYLLQPYGQGLKSKSVQVNGQPLVMVDGTLDELKPRPLRAGRTLVPVPTMGFFVW 362
Db 522 VHQYLLQPYGQGLKSKSVQVNGQPLVMVDGTLDELKPRPLRAGRTLVPVPTMGFFVW 581
Qy 363 KVNALACRYR 373
Db 582 knvnalacryr 592

RESULT 6
ID AAB81062 standard; Protein; 592 AA.
XX AAB81062;
AC AAB81062;
DT 20-JUN-2001 (first entry)
DE Human Heparanase-2 amino acid sequence.
KW Heparanase 2; human; endoglucuronidase; heparan sulphate; metastasis;
KW neangiogenesis; vaccine; autoimmune disorder; blood coagulation;
KW cancer; diabetes; ischaemia; sepsis; stroke; cardiovascular; thrombosis.
OS Homo sapiens.
XX Key Location/Qualifiers
FT Region 156..169
FT /label= Immunogenic_epitope
FT Region 249..262
FT /label= Immunogenic_epitope
FT Region 505..518
FT /label= Immunogenic_epitope
XX WO200121814-A1.
XX 29-MAR-2001.
XX 11-SEP-2000; 2000WO-EP08837.
XX 23-SEP-1999; 99EP-0118805.
PR 07-JUL-2000; 2000EP-0114649.
XX (MERE) MERCK PATENT GMBH.
XX Duecker K, Sirrenberg C;
XX WPI; 2001-308089/32.
DR N-PSDB; AAF86101.
XX New heparanase-2 polypeptide useful in diagnosing (the susceptibility
of a subject to) and as vaccines against e.g. autoimmune disorders,
cardiovascular disease, cancer, diabetes, ischemia, sepsis, stroke, or
thrombosis -
XX Claim 1; Page 42-43; 46pp; English.
XX This invention relates to a human heparanase-2 protein and the cDNA
sequence encoding it. Heparanase-2 is a member of the endoglucuronidase
family of polypeptides and it degrades heparan sulphate proteoglycans
HSPGs (ubiquitous macromolecules of cell surfaces, basement membranes and
the extracellular matrix). HSPGs support the vascular endothelium and
stabilise the structure of the capillary wall. Heparanases may be
associated with neangiogenesis and metastasis related to malignant
tumour formation. Heparanase-2 polynucleotides and proteins are useful as
vaccines for inducing an immunological response against autoimmune
disorders, blood coagulation disorders, cancer, diabetes, ischaemia,
sepsis, stroke, cardiovascular diseases, or thrombosis, as well as in
diagnosing (the susceptibility of a subject to) these diseases.
Heparanase-2 fragments may be used as immunogens to produce antibodies
immunospesific to the polypeptides, and to identify membrane bound
soluble receptors, agonists or antagonists that compete with the binding
of the polypeptide to the receptors. An antibody specific for
heparanase-2 can be used in the diagnosis of the above diseases and in

CC isolating or identifying clones expressing heparanase-2. The present
CC sequence represents heparanase-2. Three regions of heparanase-2 with high
CC immunogenicity (immunogenic epitopes) can be used to raise antibodies
CC against heparanase-2.
XX Sequence 592 AA;
Qy 1 KGCKIAQHPDVMLELQREKAAQMHVLVLLKEQFSNTYSNLILT----- 42
Db 162 kgckiaqhpdvmlvlgrekaqmhvlvllkeqfsntysnliltcarsldklynfadcsghl 221
Qy 43 -----EPNNYRTMHGRAVNGSOLGK 62
Db 222 ifalnalnrrnpunssalsllkysaskkynlswelgneppnyrtmhgravnsglqk 281
Qy 63 DYIQLKSLLPQIRIYSRASLYGNIGRPRKNVIALLDGFMKVAGSTVDVAVTWHQCYIDGR 122
Db 282 dyiqklsllqpiriysraslygnigrprknvialldgfmkvagstvdaavtwhqcyidgr 341
Qy 123 VVKVMDFLKTRLLDLSQIRIKIQVWNTYTPGKKIWLEGVVTTTSGGTNNISDSYAAAGF 182
Db 342 vvkvmdfliktrlldtldsqirkiqkvvntytpgkkiwlegvvtttsaggtnnlstdsyaagf 401
Qy 183 LWLNTGLMLANGIDVIRHSEFDHGYNHLVDONENPLPDYWLSLLYKRLIGPKVLAVHV 242
Db 402 lwntlgmlangldvvrhsfddhgynhlvdqnlfpdywlslylkrllgpkvlavhv 461
Qy 243 AGLQRKPRGRVIRDKLRIYAHCTNHHNHYVGRSTTFLIINLHRSRKKIKLAGTLRDKL 302
Db 462 aglqrkprgrvirdklriyahctnhhnhmyvrgsitfiinlhrrskkiklagtlrldkl 521
Qy 303 VHQYLLQPYGQGLKSKSVOLNGQPLVMVDGTLDELKPRPLRAGRTLVPVPTMGFFVW 362
Db 522 VHQYLLQPYGQGLKSKSVOLNGQPLVMVDGTLDELKPRPLRAGRTLVPVPTMGFFVW 581
Qy 363 KVNALACRYR 373
Db 582 knvnalacryr 592

RESULT 7
AAU07423
ID AAU07423 standard; Protein; 439 AA.
XX AAU07423;
AC AAU07423;
DT 18-DEC-2001 (first entry)
XX Human heparanase-like protein.
DE Human heparanase-like protein.
XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW antiproliferative; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;
KW nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;
XX wound healing; food additive; heparanase.
OS Homo sapiens.
XX WO200179253-A1.
PN 25-OCT-2001.
XX 11-APR-2001; 2001WO-US11643.
PF 18-APR-2000; 2000US-198123P.

XX	FA	(HUMA-) HUMAN GENOME SCI INC.	AAU07418;	AC
XX	PI	Fiscella M, Shi Y, Ebner R, Ruben SM;	18-DEC-2001 (first entry)	XX
XX	PI		Novel human extracellular matrix (ECM) protein #1.	XX
XX	DR	WPI; 2001-611720/70.		XX
XX	PT	New nucleic acids encoding extracellular matrix polypeptides, for	Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;	KW
PT	PT	diagnosing, treating, preventing or ameliorating human disorders and	antiproliferative; cardiant; vasotropic; cerebroprotective; nootropic;	KW
PT	PT	disease, such as, autoimmune, hyperproliferative or cardiovascular	neuroprotective; antibacterial; virucide; fungicide; ophthalmological;	KW
PT	PT	disorders -	extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;	KW
XX	XX		hyperproliferative disorder; neoplasm; cardiovascular disorder;	KW
XX	PS	Disclosure; Page 13-14; 308pp; English.	cardiac arrest; cerebrovascular disorder; cerebral ischemia; infection;	KW
XX	XX		nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;	KW
XX	XX		wound healing; food additive.	KW
XX	OS		Homo sapiens.	XX
CC	CC	The invention relates to novel isolated polynucleotides (I) encoding	WO200179253-A1.	XX
CC	CC	extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by	25-OCT-2001.	XX
CC	CC	(I) are used to prevent, treat or ameliorate a medical condition in e.g.	11-APR-2001; 2001WO-US11643.	XX
CC	CC	humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They	18-APR-2000; 2000US-198123P.	XX
CC	CC	are also used in diagnosing a pathological condition or susceptibility to	(HUMA-) HUMAN GENOME SCI INC.	XX
CC	CC	a pathological condition. The antibodies to the polypeptides can also be	Fiscella M, Shi Y, Ebner R, Ruben SM;	XX
CC	CC	used in alleviating symptoms associated with the disorders and in	WPI; 2001-611720/70.	XX
CC	CC	diagnostic immunoassays e.g. radioimmunoassays or enzyme linked	N-PSDB; AAS13843.	XX
CC	CC	include autoimmune diseases e.g. rheumatoid arthritis,	New nucleic acids encoding extracellular matrix polypeptides, for	XX
CC	CC	hyperproliferative disorders e.g. cardiac arrest, cerebrovascular disorders	diagnosing, treating, preventing or ameliorating human disorders and	XX
CC	CC	e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.	disease, such as, autoimmune, hyperproliferative or cardiovascular	XX
CC	CC	Alzheimer's disease, infections caused by bacteria, viruses and fungi	disorders -	XX
CC	CC	and ocular disorders e.g. corneal infection. The polypeptides can also		XX
CC	CC	be used to aid wound healing and epithelial cell proliferation, to		XX
CC	CC	prevent skin aging due to sunburn, to maintain organs before		XX
CC	CC	transplantation, for supporting cell culture of primary tissues, to		XX
CC	CC	regenerate tissues and in chemotaxis. The polypeptides can also be used		XX
CC	CC	as a food additive or preservative to increase or decrease storage		XX
CC	CC	capabilities. The present sequence represents the amino acid sequence		XX
CC	CC	of human heparanase-like protein.		XX
XX	XX		Claim 1; Page 292-293; 308pp; English.	XX
XX	XX		The invention relates to novel isolated polynucleotides (I) encoding	XX
XX	XX		extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by	XX
XX	XX		(I) are used to prevent, treat or ameliorate a medical condition in e.g.	XX
XX	XX		humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They	XX
XX	XX		are also used in diagnosing a pathological condition or susceptibility to	XX
XX	XX		a pathological condition. The antibodies to the polypeptides can also be	XX
XX	XX		used in alleviating symptoms associated with the disorders and in	XX
XX	XX		diagnostic immunoassays e.g. radioimmunoassays or enzyme linked	XX
XX	XX		include autoimmune diseases e.g. rheumatoid arthritis,	XX
XX	XX		hyperproliferative disorders e.g. cardiac arrest, cerebrovascular disorders	XX
XX	XX		e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.	XX
XX	XX		Alzheimer's disease, infections caused by bacteria, viruses and fungi	XX
XX	XX		and ocular disorders e.g. corneal infection. The polypeptides can also	XX
XX	XX		be used to aid wound healing and epithelial cell proliferation, to	XX
XX	XX		prevent skin aging due to sunburn, to maintain organs before	XX
XX	XX		transplantation, for supporting cell culture of primary tissues, to	XX
XX	XX		regenerate tissues and in chemotaxis. The polypeptides can also be used	XX
XX	XX		as a food additive or preservative to increase or decrease storage	XX
XX	XX		capabilities. The present sequence represents the amino acid sequence	XX
XX	XX		of human heparanase-like protein.	XX
XX	XX		Sequence 439 AA;	XX
XX	XX		Query Match 89.1%; Score 1743; DB 22; Length 439;	XX
XX	XX		Best Local Similarity 100.0%; Pred. No. 7.6e-177;	XX
XX	XX		Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	XX
QY	43	EPNNYTMHGRAVNSQLGKDYQLKSLQPIRYISRASLYGNIGRPRKNVIALLDGPM 102		QY
Db	109	epnnyrtmhgravnsglqgkdyqlkslqpiriyraslygnigrprknvialldgfm 168		Db
QY	103	KVAGSTVDVATWQHCVIDGRVVKVMDFLKTRLLDLSDDQIRKIQVYNTYTFGKKIWLEG 162		QY
Db	169	kvagstvdavtqhcyidgrvvkvmdflktrlldtldsdqirkikqvntytgpkilwleg 228		Db
QY	163	VVTSAGGTNNLSDSYAAGFLWLTGLMLANQIDVVIHSPFDHGCYNHLVDQNFPLPD 222		QY
Db	229	vvttsaggtnnlstdsyaaagflwltngmlanggidvvirhsffdhgynhlvdqnfplpd 288		Db
QY	223	YWLSLYKRLIGPKVLAVHVGLOKPRGRVIRDKRLRIYAHCTNNHHNHYRGSTITLFI 282		QY
Db	289	ywlslykrligpkvlavhvaglrkprgvirdklriyahctnnhhnhyrgsiltfi 348		Db
QY	283	INLHRSKKTKLAGTURDKLVHQLYLPQYQEGELKSKSVOLNQOPLVMVDDGTLPCLKPR 342		QY
Db	349	inlhrrskkiklagtlrkdvlvhyllqpyqegglkksvqlnggplvmvddgtlpeklpr 408		Db
QY	343	PLRAGRTLVIPVTMGFFVVKVNNALACRYR 373		QY
Db	409	plagrctlivppvumgffvvkvnvalacryr 439		Db
RESULT	8			
AAU07418				
ID	AAU07418	standard; Protein; 480 AA.		
XX	XX			XX

QY 103 KVGSTVDVAVTWHQCYIDGRVVKVMDFLKTRLLDLSQIRKIQKVVNTYTPGKKIWLGG 162
Db 210 kvagstvdavtwqhcylidgrvkvmdflktrlldtsdqirkiqkvntypgkklwleg 269
QY 163 VVTSAGTNNLSDSYAAGFLWNLTLGMLANOGIDVIRHSHFFDGHYNHLVDONENPLPD 222
Db 270 vvttsaggtnnlsdsyaagflwntlgmlanogidvirhshffdhgynhlvdqnfplpd 329
QY 223 YWLSLLYKRLIGPKVLAVHVAGLQRPGRVIRDKLRIYAHCTNNHNNHYRGSTITLFI 282
Db 330 ywlslllykrligpkvlavhvaglrkprgvrirdklriyahctnnhnnhyrgsitiifi 389
QY 283 INLHRSRKKIKLAGTLRDKLVHQLLOPYGQGLSKSVQLNGQPLVWVDDGTLPKLKPR 342
Db 390 inlhrrskkiklagtlrdklvhqyllqpygqglksksvqlngqplvmvddgtlpelkpr 449
QY 343 PLRAGRTLVIPTVPTMGFFVVKVNNALACRYR 373
Db 450 plragrtlviptvptmgffvkvnnalacryr 480

RESULT 9
AAB85217
ID AAB85217 standard; Protein; 480 AA.
AC AAB85217;
XX
DT 07-SEP-2001 (first entry)
XX
DE Heparanase-like protein Hpa2 splice variant #3.
XX
KW Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;
KW cytosolic; neuroprotective; cerebroprotective; immunosuppressive;
KW antiproliferative; neurotropic; antiinflammatory; antiarthritic; antiasthmatic;
KW antidiabetic; antiarteriosclerotic; vulnerary.
XX
OS Homo sapiens.
XX
PN WO200146392-A2.
XX
PD 28-JUN-2001.
XX
PF 21-DEC-2000; 2000WO-GB04963.
XX
PR 22-DEC-1999; 99GB-0030392.
PR 07-APR-2000; 2000GB-0008713.
XX
PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI McKenzie EA, Stamps AC, Terrett JA, Tyson KL;
XX WPI; 2001-418056/44.
DR N-PSDB; AAB22673.
XX
PT Novel homologs of heparanase, present in three splice variants, useful
PT for identifying agents that modulate heparanase, useful in the
PT treatment and/or prophylaxis of abnormal levels of heparanase
XX
PS Claim 1; Fig 3; 97pp; English.
XX

CC The invention provides a homologue to heparanase which is present in
CC three splice variants. The heparanase homologue polypeptide is useful in
CC the treatment of a human or non-human animal or for use in diagnosis.
CC Vectors comprising the heparanase homologue polynucleotides are useful in
CC the transformation or transfection of a prokaryotic or eukaryotic host.
CC The modulators of the polypeptide are useful in the manufacture of a
CC medicament for the treatment and/or prophylaxis of a condition/disease
CC associated with abnormal levels of the heparanase homologue, including
CC cancer, central nervous system (CNS) and neurodegenerative diseases,
CC cardiovascular diseases such as stenosis following angioplasty and
CC atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus,
CC allografts, inflammatory diseases, arthritis, vascular stenosis,
CC tumour growth and progression, asthma, Alzheimer's disease, diabetic

CC retinopathy, wound healing and inflammation. The polypeptide is also
CC useful in diagnosis and research. The present sequence represents the
CC amino acid sequence of the smallest splice variant of the heparanase-
CC like protein Hpa2 of the invention.

XX Sequence 480 AA;

Query Match 89.1%; Score 1743; DB 22; Length 480;
Best Local Similarity 100.0%; Pred. No. 8.7e-177;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 EPNVYRTMHGRAVNSQLGKDYIQLKSLQPIRIYSRASLYGNIGRPRKNNVIALDGM 102
Db 150 epnyvrtmhgravnsqigkdyiqkllqpiriyraslygnigrprknnvialldgfm 209
QY 103 KVGSTVDVAVTWHQCYIDGRVVKVMDFLKTRLLDLSQIRKIQKVVNTYTPGKKIWLGG 162
Db 210 kvagstvdavtwqhcylidgrvkvmdflktrlldtsdqirkiqkvntypgkklwleg 269
QY 163 VVTSAGTNNLSDSYAAGFLWNLTLGMLANOGIDVIRHSHFFDGHYNHLVDONENPLPD 222
Db 270 vvttsaggtnnlsdsyaagflwntlgmlanogidvirhshffdhgynhlvdqnfplpd 329
QY 223 YWLSLLYKRLIGPKVLAVHVAGLQRPGRVIRDKLRIYAHCTNNHNNHYRGSTITLFI 282
Db 330 ywlslllykrligpkvlavhvaglrkprgvrirdklriyahctnnhnnhyrgsitiifi 389
QY 283 INLHRSRKKIKLAGTLRDKLVHQLLOPYGQGLSKSVQLNGQPLVWVDDGTLPKLKPR 342
Db 390 inlhrrskkiklagtlrdklvhqyllqpygqglksksvqlngqplvmvddgtlpelkpr 449
QY 343 PLRAGRTLVIPTVPTMGFFVVKVNNALACRYR 373
Db 450 plragrtlviptvptmgffvkvnnalacryr 480

RESULT 10
AAY97634
ID AAY97634 standard; Protein; 480 AA.
XX
AC AAY97634;
XX
DT 20-APR-2001 (first entry)
XX
DE Human heparanase, hnhp1 pn5 form, protein sequence.
XX
KW Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;
KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
KW gene therapy; human.
XX
OS Homo sapiens.
XX
PN WO200100643-A2.
XX
PD 04-JAN-2001.
XX
PF 19-JUN-2000; 2000WO-IL00358.
XX
PR 25-JUN-1999; 99US-0140801.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI Pecker I, Michal I, Itzhaki H;
XX WPI; 2001-137930/14.
DR N-PSDB; AAA91099.
XX
PT New polynucleotides and polypeptides that are distantly homologous to
PT heparanase, useful in wound healing, as well as in gene therapy
PT protocols for angiogenesis, restenosis, atherosclerosis, or
PT inflammation -

```
XX PS Claim 10; Page 63; 67pp; English.
XX
CC This sequence represents a heparanase of the invention.
CC The heparanase DNA and protein sequences are useful in wound healing,
CC angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary
CC diseases, neurodegenerative diseases (such as Sclerose, Alzheimer's
CC disease, and Creutzfeldt-Jakob disease) or viral infections. The
CC heparanase coding sequence is particularly useful in gene therapy.
XX
XX Sequence 480 AA:
XX
Query Match 89.1%; Score 1743; DB 22; Length 480;
Best Local Similarity 100.0%; Pred. No. 8.7e-177; Indels 0; Gaps 0;
Matches 331; Conservative 0; Mismatches 0;
Qy 43 EPNNYRTMHGRAVNGSQLGKDYIQLKSLQPIRIYSRASLYGPNIGRPRKNVIALLDGFM 102
Db 150 epnnyrtnhgravnsgqlgkdyiqklsllqpriyrstaslygpnigrprknvialldgfm 209
Qy 103 KVAAGTVDATWQHGYIDGRVVKVWDFLKLRLDLSQIRKIQKVNTYTPGKKIWLEG 162
Db 210 kvagstvdavtqhcyidgrvkvkvmfklrlldtsdqirklgkvntytpgkklwleg 269
Qy 163 VVTSAGCTNNLSDSYAAGFLWNTLGLMANGIDVIRHSHFFDHGYNHLVDQNFNPLPD 222
Db 270 vvttsagctnnlssdsyagflwntlglmanqgidvirhsffdhgynhlvdqnfnpdp 329
Qy 223 YWLSLLYKRLIGPKVLAVHVGAGLQRPGRVIRDKLRIYAHCTNNHNNHNVYRGSITLFI 282
Db 330 ywlslllykrligpkvlavhvaglgkprgprvirdklriyahctnnhnnhnyvrgsitlfi 389
Qy 283 INLHRSRKIKLAGTLRDLKLVHQLLQPYGQGLKSKSVQLNGQPLVMVMDGTLPKLKPR 342
Db 390 inlhrsrrkiklagtlrdklvhlqyilqpygqglksksvqlngqplvmvmdgtlpelkpr 449
Qy 343 PLRAGRTLVIPVPTMGFFVKNVNALACRYR 373
Db 450 plragrtlvippvtmgffvknvnalacryr 480
RESULT 11
AA97633
ID AA97633 standard; Protein; 538 AA.
XX
AC AA97633;
XX
XX 20-APR-2001 (first entry)
XX
XX Human heparanase, hnhp1 pn9 form, protein sequence.
XX
XX Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Sclerose;
XX atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
XX neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
XX gene therapy; human.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX
XX FT Misc-difference 305
XX /note= "encoded by GAC"
XX
XX WO200100643-A2.
XX
XX 04-JAN-2001.
XX
XX 19-JUN-2000; 2000WO-IL00358.
XX
XX 25-JUN-1999; 99US-0140801.
XX
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
```

```
PI Pecker I, Michal I, Itzhaki H;
XX
DR WPI; 2001-137930/14.
DR N-PSDB; AAA91098.
XX
XX New polynucleotides and polypeptides that are distantly homologous to
XX heparanase, useful in wound healing, as well as in gene therapy
XX protocols for angiogenesis, restenosis, atherosclerosis, or
XX inflammation -
XX
XX Claim 10; Page 61-62; 67pp; English.
XX
XX This sequence represents a heparanase of the invention.
XX The heparanase DNA and protein sequences are useful in wound healing,
XX angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary
XX diseases, neurodegenerative diseases (such as Sclerose, Alzheimer's
XX disease, and Creutzfeldt-Jakob disease) or viral infections. The
XX heparanase coding sequence is particularly useful in gene therapy.
XX
XX Sequence 538 AA:
XX
Query Match 89.0%; Score 1740; DB 22; Length 538;
Best Local Similarity 96.0%; Pred. No. 2.2e-176; Indels 2; Gaps 1;
Matches 337; Conservative 1; Mismatches 11;
Qy 25 LVLLKEQFSNTY--SNLILTEPNNYRTMHGRAVNGSQLGKDYIQLKSLQPIRIYSRASL 82
Db 188 lsllkysaskkyniswelgnepnnrtnhgravnsgqlgkdyiqklsllqpriyrstas 247
Qy 83 YGPNIGRPRKNVIALLDGFMKVGSTVDATWQHGYIDGRVVKVWDFLKLRLDLSQI 142
Db 248 ygpnlgrprknvialldgfmkvagstvdavtqhcyidgrvkvkvmfklrlldtsaqi 307
Qy 143 RKIQKVNTYTPGKKIWLEGVVTTSAGCTNNLSDSYAAGFLWNTLGLMANGIDVIRH 202
Db 308 rkikqvntytpgkklwlegvvttsagctnnlssdsyagflwntlglmanqgidvirh 367
Qy 203 SFFDHGYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHVGAGLQRPGRVIRDKLRIY 262
Db 368 sffdhgynhlvdqnfnpdpylwlslllykrligpkvlavhvaglgkprgprvirdklriy 427
Qy 263 AHCTNNHNNHNVYRGSITLFIINLHRSRKIKLAGTLRDLKLVHQLLQPYGQGLKSKSVQ 322
Db 428 ahctnnhnnhnyvrgsitlfiinlhrsrrkiklagtlrdklvhlqyilqpygqglksksvq 487
Qy 323 LINGQPLVMVMDGTLPKLKPRPLRAGRTLVIPVPTMGFFVKNVNALACRYR 373
Db 488 lngqplvmvmdgtlpelkprplragrtlvippvtmgffvknvnalacryr 538
RESULT 12
AA99905
ID AA99905 standard; Protein; 214 AA.
XX
AC AA99905;
XX
XX 07-JAN-2002 (first entry)
XX
XX Human excretory related polypeptide SEQ ID NO 642.
XX
XX Human; neutrotropic; neuroprotective; cytostatic; dermatological; virucide;
XX immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnary;
XX antiparkinsonian; antisklking; antianaemic; antiarthritis; cancer;
XX antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
XX antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;
XX antiparasitic; cardiant; immune disorder; cardiovascular disorder;
XX neurological disease; infection; nephrotropic; gene therapy; vaccine;
XX excretory system.
XX
XX Homo sapiens.
XX
XX WO200155313-A2.
XX
XX
```


XX PD 02-AUG-2001.
XX XX 17-JAN-2001; 2001WO-US01323.
XX PF 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA Rosen CA, Barash SC, Ruben SM;
XX PI
XX XX

DR WPI; 2001-465569/50.
DR N-PSDB; AA198878.
XX
XX Isolated nucleic acid molecule encoding excretory system antigen is
PT used in preventing, treating or ameliorating a medical condition -
PT
XX
PS Claim 11; SEQ ID NO 642; 574pp + Sequence Listing; English.
XX
XX The invention relates to novel excretory system related human
CC polynucleotides (AA198567-AA199503) and the encoded proteins
CC (AAK99594-AAK99913) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy, especially
CC disorders related to the excretory system. The genes are isolated
CC from a range of human tissues disclosed in the specification. The
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: the sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 214 AA:
SQ

Query Match 56.9%; Score 1112; DB 22; Length 214;
Best Local Similarity 100.0%; Pred. No. 5.1e-110; Indels 0; Gaps 0;
Matches 210; Conservative 0; Mismatches 0;

QY 164 VTTSGAGTNNLSDYAAAGFLWLTGMLANQGDVIRHSFFDHGYNHLVDQNFNPLPDY 223
DB 5 vttsaggtnnlstdyaaagflwltgmlanqgdvirsffdhgynhlvdqnfnpdpdy 64

QY 224 WLSLLYKRLIGPKVLAVHVGAGLQKRPGRVIRDKLRIYAHCTNNHNNHYVRSITLFI 283
DB 65 wlslllykrligpkvlavhvgaglrqkprgrvirrdklriyahctnnhnnhyvrsitlfi 124

QY 284 NLHRSRKKIKLACTLRDKLVHVLQPYGOEGLKSKSVOLNGOPLVMVDDGTLPKLRP 343
DB 125 nlhrsrrkkiklactlrdklvhvlqpygoegllksksvqlngoplvmdvddgtlpelkprp 184

QY 344 LRAGRTLVIPTVTMGFFVVKNNALACRYR 373
DB 185 lragrtlviptvtmgffvkvnnalacryr 214

RESULT 13
AAK43704
ID AAK43704 standard; Protein; 214 AA.
XX
XX AAK43704;
XX
XX 24-OCT-2001 (first entry)
DT
DE Human bladder antigen, SEQ ID NO: 98.
XX
XX Human; bladder antigen; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antiulcer; vulnerary; anticonvulsant; antibacterial;
KW antifungal; antiparasitic; cardiant; gene therapy; cancer;
KW immune disorder; cardiovascular disorder; wound healing; infection;
KW neurological disease.
XX
XX Homo sapiens.
OS
XX WO200159064-A2.
PN

Qy	256	RDKLRIYAHCTNNHHNHYVGRSGTFLFINLHRSEKKIKIAGTLRDKLVHQYLLLPVQBG	315
Db	430	--klrvylctctndprykegdltiyainlmwtkyrlpypfsnqvdqyilrlpgph	487
Qy	316	LKSKSVQLNGPLVMVDGTLPELKPRLRAGRTLPIPPVTMGFFVVKNNNALAC	370
Db	488	llksvqlngltlkmvddqtlpplmekpirlpgssslgpafsyffvrlrnakvaac	542

Search completed: July 30, 2002, 08:14:04
Job time: 341 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:16:32 ; Search time 71.68 Seconds
(without alignments)
500.018 Million cell updates/sec

Title: us-09-836-461-2_copy_162_534

Perfect score: 1956

Sequence: 1 KCKKTAQHDPVLMLEQREKA.....PVTMGFFVKNVNALACRYR 373

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1743	89.1	480	2 JC7506	heparanase protein
2	301	15.4	521	2 T45608	hypothetical prote
3	155	7.9	190	2 T01953	hypothetical prote
4	102.5	5.2	1260	2 T14022	reverse transcript
5	101	5.2	535	2 S58740	cytochrome-c oxida
6	100.5	5.1	630	2 C71374	probable glucose i
7	95	4.9	512	2 T11261	cytochrome-c oxida
8	94.5	4.8	460	2 G90554	p46-like (mycoplas
9	94	4.8	625	2 G86855	glucose inhibited
10	94	4.8	1517	2 B81393	DNA-directed RNA p
11	93	4.8	320	2 D90478	conserved hypothet
12	92.5	4.7	328	2 AC1210	drdp-D-glucose 4,6
13	92.5	4.7	876	2 G89952	DNA polymerase I
14	92.5	4.7	2241	2 S09811	hypothetical prote
15	92	4.7	461	1 S65187	GPI-anchor biosynt
16	92	4.7	534	2 S17993	cytochrome-c oxida
17	91.5	4.7	335	2 E84992	tryptophan--trna l
18	90	4.6	650	2 F91654	regulatory protein
19	90	4.6	709	2 T16584	hypothetical prote
20	90	4.6	1616	2 T16600	vitellogenin vit-1
21	89.5	4.6	898	2 B8471	cytoplasmic aconit
22	88.5	4.5	286	2 A86576	chromosome partiti
23	88.5	4.5	286	2 B72047	chromosome partion
24	88.5	4.5	341	2 D70761	probable fatty-acy
25	88.5	4.5	456	2 T24442	hypothetical prote
26	88.5	4.5	1272	2 C90593	hypothetical prote
27	88	4.5	555	2 AE2147	cytochrome c oxida
28	88	4.5	1106	1 CHRBA2	calcium channel pr
29	87.5	4.5	527	2 S44081	lls globulin - oat

30	87.5	4.5	534	2 JU0148	cytochrome-c oxida
31	87.5	4.5	557	1 ODNCL	cytochrome-c oxida
32	87.5	4.5	627	2 A69663	DNA mismatch repai
33	87.5	4.5	770	2 S77523	hypothetical prote
34	87.5	4.5	820	2 G75251	cytochrome-c oxida
35	87.5	4.5	828	2 S56250	probable membrane
36	87	4.4	537	2 F86938	probable methionyl
37	87	4.4	3194	2 D71917	toxin-like outer m
38	86.5	4.4	494	2 D84860	probable protein k
39	86.5	4.4	534	2 S78640	cytochrome-c oxida
40	86.5	4.4	1009	2 E89792	hypothetical prote
41	86.5	4.4	1421	2 T49500	hypothetical prote
42	86	4.4	805	2 E97717	virB4 protein prec
43	86	4.4	1315	2 T03167	hypothetical prote
44	86	4.4	1562	2 T29146	hypothetical prote
45	85.5	4.4	429	1 T38146	dihydrofolate redu

RESULT 1

JC7506

heparanase protein 2a - human

C:Species: Homo sapiens (man)

C:Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 01-Dec-2000

C:Accession: JC7506

R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hirccock, M.; Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000

A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase famli

A:Reference number: JC7506

A:Accession: JC7506

A:Molecule type: mRNA

A:Residues: 1-480 <MCK>

A:Cross-references: GB:AF282885

C:Comment: This protein, a intracellular membrane-bound enzyme, has biological and th

C:Genetics:

A:Gene: hpa2a

A:Map position: 10q23-10q24

C:Keywords: heparin binding; membrane bound

Query Match 89.1%; Score 1743; DB 2; Length 480;

Best Local Similarity 100.0%; Pred. No. 4.8e-138;

Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ALIGNMENTS

RESULT 1

JC7506

heparanase protein 2a - human

C:Species: Homo sapiens (man)

C:Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 01-Dec-2000

C:Accession: JC7506

R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hirccock, M.; Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000

A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase famli

A:Reference number: JC7506

A:Accession: JC7506

A:Molecule type: mRNA

A:Residues: 1-480 <MCK>

A:Cross-references: GB:AF282885

C:Comment: This protein, a intracellular membrane-bound enzyme, has biological and th

C:Genetics:

A:Gene: hpa2a

A:Map position: 10q23-10q24

C:Keywords: heparin binding; membrane bound

Query Match 89.1%; Score 1743; DB 2; Length 480;

Best Local Similarity 100.0%; Pred. No. 4.8e-138;

Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 EPNNRTMHGRAVNSQLGKDYIQLKSLLOPIRIYSRASLYGNPIGRPKRKNVIALLDGFM 102

Db 150 EPNNRTMHGRAVNSQLGKDYIQLKSLLOPIRIYSRASLYGNPIGRPKRKNVIALLDGFM 209

QY 103 KVAGSTVDATVWQHCVYIDGRVVKVMDFLKTRLLDLSDDIRKIQKVVNTYTPGKKIWLSE 162

Db 210 KVAGSTVDATVWQHCVYIDGRVVKVMDFLKTRLLDLSDDIRKIQKVVNTYTPGKKIWLSE 269

QY 163 VVTSAGGTNNLSDSYAAGFLWLTGLMLANQIDVVIHRHSFFDHGYNHLVDQNFPLPD 222

Db 270 VVTSAGGTNNLSDSYAAGFLWLTGLMLANQIDVVIHRHSFFDHGYNHLVDQNFPLPD 329

QY 223 YWLSLLYKRLIGPKVLAVHAGVAGLQRPGRVIRQDKLRIYAHCTNHHNNHYRGSTITLFI 282

Db 330 YWLSLLYKRLIGPKVLAVHAGVAGLQRPGRVIRQDKLRIYAHCTNHHNNHYRGSTITLFI 389

QY 283 INLHRSRKKIKLAGTLRDKLVHQLLOPYQBELSKSVQLANGQPLVMDVDDGTLPKLKPR 342

Db 390 INLHRSRKKIKLAGTLRDKLVHQLLOPYQBELSKSVQLANGQPLVMDVDDGTLPKLKPR 449

QY 343 PLRAGRTLVIPVMTGFFVYVKNVNALACRYR 373

Db 450 PLRAGRTLVIPVMTGFFVYVKNVNALACRYR 480

RESULT 2

QY 367 ALA 369
Db 705 ALA 707

RESULT 5

S58740

Cytochrome-c oxidase (EC 1.9.3.1) chain I - yeast (Hansenula wingei) mitochondrion

C:Species: mitochondrion Hansenula wingei

C:Date: 19-Mar-1997 #sequence_revision 09-May-1997 #text_change 07-Dec-1999

C:Accession: S58740

R:Seikito, T.; Okamoto, K.; Kitano, H.; Yoshida, K.

Curr. Genet. 28, 39-53, 1995

A:Title: The complete mitochondrial DNA sequence of Hansenula wingei reveals new charact

A:Reference number: S58740; MUID:96022424

A:Accession: S58740

A:Molecule type: DNA

A:Residues: 1-535 <SEK>

A:Cross-references: EMBL:D31785

A:Note: the authors translated the codon CAA for residue 54 as Gly

C:Genetics:

A:Gene: cox1

A:Genome: mitochondrion

A:Genetic code: SGC3

A:Introns: 241/3

C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
ductase; respiratory chain; transmembrane protein

F:11-458/Domain: cytochrome-c oxidase chain I homology <COI>

F:63,379/Binding site: heme a iron (His) (axial ligands) #status predicted

F:242,291-292/Binding site: copper (His) #status predicted

F:242-246/cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted

F:246/Binding site: oxygen (Tyr) #status predicted

F:320/Binding site: myristate (lys) (covalent) #status predicted

F:369/Binding site: magnesium (His) (shared with chain II) #status predicted

F:377/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match

Best Local Similarity 5.2%; Score 101; DB 2; Length 535;

Matches 68; Conservative 46; Mismatches 102; Indels 90; Gaps 19;

QY 74 IRIYSRASLYGNTPGRKKNVIALLDGFMKVGSTVDATW-OHCYIDGRVVKVMDFLKT 132

Db 260 VSTYSKPKVFG-----ISMVYAMASIA--FLGLVMSHHMYIVG-----LD-ADT 302

QY 133 RLDTLDPQIRKIKQVNTYTPGKKIWLEGVVTTSAGTNNLSDS--YAAGFLWLNTLG- 189

Db 303 RAYFSTSTWVAVPTGKIFPS-----WL-----ATLYGGSIRLAVPMLYAFLLFTIGG 353

QY 190 ----MLANOGIDVVIRHSFFDHGYNHLY---DQNFNPLPDYV-----LSLLYKR----- 231

Db 354 LTGVALANASLDVAFHDYVYVGHFYVLSMGAIFSLFAGYVYWSQPIGLGLYFNERLAQI 413

QY 232 -----LIGPKV--LAVHVAGLQRPGRVIRDKLRIYAHCNTNHNHNYVRG----- 276

Db 414 QFWLIFGANVIFMPHFLGLQGMPR-----RIPDYPDAYAGWNYVSSIGSVIAII 464

QY 277 STTLFIINLHRSRKIKLAGTLRDLKHVQ---YLLQP-----YGOGLKSKSQV--L 323

Db 465 SLALFIYIID-----QLINGLTNKNIDKNKSVVYSKAPDFVESNTIFANNSTKSASIEFL 519

QY 324 NGQPLV 329

Db 520 NSPPAI 525

RESULT 6

C71374

probable glucose inhibited division protein A (glda) - syphilis spirochete

C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 26-Aug-1999

C:Accession: C71374

R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998

A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A:Reference number: A71250; MUID:98332770

A:Accession: C71374

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-630 <COL>

A:Cross-references: GB:AE001189; GB:AE000520; NID:g3322293; PIDN:AAC65038.1; PID:g332

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0044

C:Superfamily: glda protein

Query Match 5.1%; Score 100.5; DB 2; Length 630;

Best Local Similarity 21.3%; Pred. No. 2.8;

Matches 77; Conservative 52; Mismatches 129; Indels 103; Gaps 20;

QY 57 GSOLGK--DYIQLKSLLOPIRIYSRASLYGNTPGRKKNVIALLDGFMKVGSTVDATW 114

Db 68 GGEMGFADACMIQ-----YRLNLSR--GPAVQAPRIQADKFL--YAKVKYTILECTQH 118

QY 115 QHCYIDGRVVKVMDFLKTRLLDTLSQIRKIQVNTYTPGKKIWLEGVVTTSAGTNNL 174

Db 119 LHLYQD---TVVDVCSNTTDAGYVAYGAHAVVT--ARGRRISARAVLTGT----- 166

QY 175 SDSYAAAGFLWL-----NTLGLANOGIDVVIRHSFFDHG-----YNHLVDQNF 217

Db 167 --TFMEGRVYIGEYEAPEGRLGEHAAELGAAALRGKGFQMGRLKGTGPVRLRKSV----- 221

QY 218 NPLPDYWLSLYKLR---LIGPKVLAVHVGAGLQRPGRVIRDKLRIYAHCNTNHNHNY 273

Db 222 -----LSVMEKQCADAIMRPFSEA-HV-----EINRPHADCYINTNERTHOL 263

QY 274 VSGSTLFIINLHRS---RKIKLAGT-----LRDLKHVQYLLQPYGOGLKS 318

Db 264 IRE-----NPHRSPPFSGRKAVGTVCPSIEDKVRKFPDRIRHQLVIEP---EGLDT 313

QY 319 KSVQLNGPLVMVD-----GTLPELK-----PRLRAGRTLVIPTVMGF-FVVKVNA 367

Db 314 EELYINGLSCLPEDIQDEMIRTIFGMRVAVITPAYAVDYAVLFPVLGIDLQTKRVSG 373

QY 368 L 368

Db 374 L 374

RESULT 7

T11261

Cytochrome-c oxidase (EC 1.9.3.1) chain I - hardbacked tick (Ixodes hexagonus) mitoch

C:Species: mitochondrion Ixodes hexagonus

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000

C:Accession: T11261

R:Black IV, W.C.; Roehrdanz, R.L.

Mol. Biol. Evol. 15, 1772-1785, 1998

A:Title: Mitochondrial gene order is not conserved in arthropods: prostrate and meta

A:Reference number: 217252; MUID:99083443

A:Accession: T11261

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-512 <BLA>

A:Cross-references: EMBL:AF081828; NID:g4164542; PID:g4164544; PIDN:NAD05505.1

C:Genetics:

A:Genome: mitochondrion

C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology

F:10-256/Domain: cytochrome-c oxidase chain I homology <COI>

F:60,377/Binding site: heme a iron (His) (axial ligands) #status predicted

F:239,289/Binding site: copper (His) #status predicted

F:239-243/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted

F:243/Binding site: oxygen (Tyr) #status predicted
F:367/Binding site: magnesium (His) (shared with chain II) #status predicted
F:375/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 4.8%; Score 95; DB 2; Length 512;
Best Local Similarity 24.6%; Pred. No. 6;
Matches 42; Conservative 22; Mismatches 57; Indels 50; Gaps 11;

QY 112 VTW-QHCYIDGRVVKVDFLKTRELDLTLDQIRKIQKVVNTYTPGKKIKWLEGVVTTSAGG 170
DB 285 IWAHMETVG-----MD-IDTRAYFAATWMLIAVPTGKIFS-----WL-----ATLHGS 329
QY 171 TNNLSDS--YAAGFLWLNTLG-----MLANQIDVIRHSFFDHGYNHLV----- 213
DB 330 NINFNSMLWLVGFVFLPTLGLTGILANSSIDILHDTYVVVAHFHYVLSMGAVFAIM 389
QY 214 -----DQNFNPLPDYWLSL-LYKRLIGPKV--LAVHVAGLQKPR 250
DB 390 GSITHWFFLPFGMNFNSL---WLKIQFYSMFIGNMTFFPOHFLGLSSMPR 437

RESULT 8
G90554
p46-like (mycoplasma hyorhinis) lipoprotein [imported] - Mycoplasma pulmonis (strain UAB
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: G90554
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallissou, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: G90554
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-460 <KUR>
A:Cross-references: GB:AL445566; PID:g14089757; PIDN:CAC13516.1; GSPDB:GN00153
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYP0_3430
A:Genetic code: SGC3

Query Match 4.8%; Score 94.5; DB 2; Length 460;
Best Local Similarity 19.6%; Pred. No. 5.7;
Matches 78; Conservative 53; Mismatches 145; Indels 121; Gaps 18;

QY 28 LKEQFSNTYSNLILTPEPNYRTMHGRAVNSQLGKDYIOLKSLLOPIRIYSRASLYGPN 87
DB 55 VSEKISDGSIVITVDPENQ-----WVETKKQLD---AYSKT----- 89
QY 88 GRPKNVIALLDGFMKAVAGSTVDVAVTWHQCYIDGRVVKVDFLKTRELDLTLDQ----- 141
DB 90 -----NGFEHISNHVKAQAQNSFVDAEIAKTNKPKVVLMLGAADSGNATQA 138
QY 142 -----IRRIQK-----VVNTYTPGKKIKWLEGVVTTS-----GGTNKL 174
DB 139 IESTNAAQOIAVDREIFHKISNNYVAFNNTYVGQ---LOGLALISGIYKQKQEPK 195
QY 175 SDSYAAGFLWNLV-----GMLANQIGI-DVIRHSFFDHGYNHL-----VDQNFNPLPDY 224
DB 196 TLEAKTYVANKLASEKGFVALAGAPEDNNSHLFKFGAMDVITAMKIDSNLKYFGE 255
QY 225 LSLLYKRLIGPKVLAVHAGLQKPRGRV---IRDKLRIYACTNNHNNYVRGS---I 278
DB 256 GADVFNRLSG-----DSKPTDQELTSAITSHFRVVAQA---WNWYSQKQFLD 300
QY 279 TLFITNLHRSKKIKTAGTL--RDKLVHQLLOPYGOEGLKSKSVOLNGOPLVWVDGTL 336
DB 301 TLPKSTFDSKKN-NLAVALAPNDEMAQAATISIEQGLDPKKIYITGQ-----DSNQ 353
QY 337 PELKPRPLRAGRTLIVP-----PVTMGFFVVK 364

Db 354 PSLRSIDNETGQNMITSKEDWKIAAISTALAYYIVKN 390

RESULT 9
G86855
glucose inhibited division protein Gida [imported] - Lactococcus lactis subsp. lactis
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: G86855
R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Eh
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: G86855
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-625 <STO>
A:Cross-references: GB:AE005176; PID:g12724877; PIDN:AAK05945.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: gida
C:Superfamily: gida protein

Query Match 4.8%; Score 94; DB 2; Length 625;
Best Local Similarity 19.6%; Pred. No. 9.6;
Matches 72; Conservative 58; Mismatches 119; Indels 118; Gaps 17;

QY 68 KSLLOPIRYSRASL-YCPNIGRPKKNVIAL-LDGFPMKVAGSTVDVAVTWHQCYID---GR 122
DB 31 KTLMTINLNMVAFMPCNPISGGSAKGIVVREIDALGEGMRNIDKTYIQMKMLNTGKGP 90
QY 123 VVKVM-----DFLKTRELDLTLDQ-----IRKI-----QKVVNTYTPGKKIKW 160
DB 91 AVRALRAQADKDEYADSMKNTVSDQENLTLRQGMVEELILDEEKKVIGIKTSTGYGA 150
QY 161 EGVVTT-----SAGGTNNLSDSYAAGFLWNLGMLANQIGDVIRHSF 204
DB 151 KAVITTTGTALRGRIIIGELKYSGGPNNSLS-----SIGLADN-----LREIG 193
QY 205 FDHGVNHLVDQNFNPLPDYMLSLYKRLIGPKVLAVHAGLQKPRPG----- 252
DB 194 FEIG-----REFKTGTPPRVLASSIDYDKTEIQGDEAPNHFSEMS 234
QY 253 --RVIRDKLRIYACTNNHNNYVRGSTLTFLIINLHRS---RKKIKLAG-----TLRDKL 302
DB 235 DENYKLDQIPCWLTAVTNTSTILRG-----NLHRAPLFSGIVKGVGPRYCPSIDKI 287
QY 303 V-----HQYLLQPYGOEGLKSKSVOLNGOPLVWVDGTLPELKPRLRAGRTLIVIPP 355
DB 288 TRFADKPRHQLFLEP---EGRNTEEVYIGGLSTSPEDVQDFLVKSIPLGNAQMMRPGY 344
QY 356 TMGFFV 362
DB 345 ATEYDVV 351

RESULT 10
B81393
DNA-directed RNA polymerase (EC 2.7.7.6) beta' chain C10479 [imported] - Campylobacte
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 08-Sep-2000
C:Accession: B81393
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chil
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A:Reference number: A81250; MUID:20150912
A:Accession: B81393
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1517 <PAR>
A:Cross-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB75117.1; PID:g696

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: rpoC; Cj0479

C:Superfamily: Escherichia coli DNA-directed RNA polymerase beta' chain

C:Keywords: nucleotidyltransferase

Query Match 4.8%; Score 94; DB 2; Length 1517;

Best Local Similarity 20.2%; Pred. No. 34;

Matches 68; Conservative 59; Mismatches 99; Indels 110; Gaps 20;

Qy 91 RKNVIALLD-----GFWKVGSTVDVNT---WQHCYIDGRVVKVMDFL---KPLLDLTL 138

Db 607 KKDIAALVDYVYKGGLETTASFDRKLNKLFYATKAGISISTADIIVPNDKQKAIDEA 666

Qy 139 SDQIRKIQKVNTYTPGKIKWLEGVVTTSAGTNNLSDSYAAGFTLWTLGLMANQGDIV 198

Db 667 KKQVREIQ---NSYNLG-----LITSGERYNKIID-----IWKSTNNVLSKEMMKL 709

Qy 199 VIRHSFFDHGYNHLVDQNFNLPDYWLSLLYKRLIGPKVLAVHVA-----GLQRKPRG 252

Db 710 VEKDK---EGFN-----SIYMMADSGARGSAQISQLAAMRGLTKP-DG 750

Qy 253 RVIRDKLRIYAHCTNHHNNHNVRG-SITLFIINLHRSRK-----KIKLAGTLRDLKVH 304

Db 751 SIETPI-----ISNREGUNLVLEYFISTHGARKGLADTALKTANAGAYLTRKLID 800

Qy 305 -----QYLQPYGOGLKSKSVOLNGQPLVMVDGTLPELKPRLRAGRTL---VIPPV 356

Db 801 VAQNVKITIEDCGTH---EGVEINE---ITADSSIIETLEERIL--GRVLAEDVIDPIT 851

Qy 357 -----MGFFVVKVNVN---ALACR 371

Db 852 NSVLFAEGLTMDDEKAKILGESGIKSVNIRTPITCK 887

RESULT 11

D90478

conserved hypothetical protein [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001

C:Accession: D90478

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Chan-

Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, B.

arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.

submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: D90478

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-320 <KUR>

A:Cross-references: GB:AE006641; NID:g13816361; PIDN:AAK43083.1; GSPDB:GN00155

C:Genetics:

A:Gene: SS02979

Query Match 4.8%; Score 93; DB 2; Length 320;

Best Local Similarity 23.7%; Pred. No. 4.6;

Matches 56; Conservative 31; Mismatches 77; Indels 72; Gaps 12;

Qy 23 MHLVLLKEQFSNTYSNLILTEPNNTYTMHGRAVNGSQLGKDYIOLKSLLOPIRIYSRASL 82

Db 1 MNLATIEFE-SNVLNRDNLNDPKYRRV-----GIIPKDYEG-----RPLIYLSGYL 47

Qy 83 -----YGP-----NIGRPKNVIALLDGFMKVGAGTVDVAVTQHCYIDG 121

Db 48 SSSLTQINYNPLGEDMWSKVERLSNEGKMGSVIPLDMFTKVGQ-----YINS 98

Qy 122 RVVKVW-DFLKRLDLSQITRKIQVYNTVTPCKIWLGVVTSAGGTNNLSDSYAA 180

Db 99 SAVGMYEDFLVKELIPFLKDF-KSDKI-----GIFCHSSGGYGAL-----IL 140

Qy 181 GFLWNLTLGLMANOGIDVIRHSFFDHGYNHLVDQNFNPL-----PDYWLSLLYKR 231

Db 141 GMKYPNTIKAIADHAGD-----AVFEYVYLTPFPRAIEQLRRFKTPEEWLENYWKK 191

RESULT 12

AC1210

dTDP-D-glucose 4,6-dehydratase homolog lmo1083 [imported] - Listeria monocytogenes (s

C:Species: Listeria monocytogenes

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001

C:Accession: AC1210

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec

; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Krefet, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;

Ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vasquez-Boland, J.A.; Voss, H.; Wehla

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AC1210

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-328 <GLA>

A:Cross-references: GB:NC_003210; PIDN:CAC99161.1; PID:g16410485; GSPDB:GN00177

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: lmo1083

C:Superfamily: Escherichia coli UDPglucose 4-epimerase; UDPglucose 4-epimerase homolo

Query Match 4.7%; Score 92.5; DB 2; Length 328;

Best Local Similarity 20.4%; Pred. No. 5.2;

Matches 57; Conservative 41; Mismatches 91; Indels 91; Gaps 15;

Qy 116 HCYIDGRVVKVMDFLKRLDLSQIRKIQKVVN-----TYPGKKIWLEGVVT 165

Db 53 HVFVEGNICD-YDLVKKLVTD-----HKDITIVFAAESHVDRSIINPG---IFIE---- 99

Qy 166 TSAGTNNL-----SDSYAAGFLWNT---LGLMANQG-----IDV 198

Db 100 TNVQGTLLNNVAKELNVAKYLVQVTSDEYVSLGETGYFTETPIAPNSPYSKASADL 159

Qy 199 VIRHSFFDHGYNHLV---DQNFNP---LPDYWLSLLYKRLIGPKVLAVHVGLOKRPGR 253

Db 160 LVRSYFYETGLNVNITRCSNNYPHFPFKLIPMTITNGDGENLPY-----GDGK 211

Qy 254 VIRDKLRIYAHN-----TNHHNNYVVGSTLFIINLHRSRKKIKLAG 296

Db 212 NIROWLHVSDHCAADLVHNGKSGEVNVGGHNERTNNEIHHIIVDDLNLJSKDKIVY-- 269

Qy 297 TLIRDKLHVH---QYLLQPYGOGLKSKSVOLNGQPLVMVDG 334

Db 270 -VEDRLGHDLRVAIDP-----KKIETELGWEPKYTFDTG 302

RESULT 13

GN8952

DNA polymerase I [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001

C:Accession: GN8952

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; K

ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: GN8952

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-876 <KUR>

A:Cross-references: GB:BA000018; PID:g13701486; PIDN:BAB42780.1; GSPDB:GN00149

A:Experimental source: strain N315

Tue Jul 30 10:11:58 2002

C:Genetics:
A:Gene: polA
C:Superfamily: DNA-directed DNA polymerase I

Query Match 4.7%; Score 92.5; DB 2; Length 876;
Best Local Similarity 21.7%; Pred. No. 21;
Matches 62; Conservative 51; Mismatches 104; Indels 69; Gaps 14;

Qy 22 QMHLVLLKEQFSNTYSNLI--LTPEN-----NYRTMHGRAVNSQLGKDYIQLKS 69
Db 342 EKHIIVNADDINN-YAEIVSWLENPTKKVYDAKTTIVASHRLGDIQNIQSFIMILASY 400
Qy 70 LLOPTR-----IYSRASLYGPNI-----GPRKRNVIALLDGFMKVAGSVDAVTWQHC 117
Db 401 IIDPRTTSDVQSVVSLYGQSFVKDDVSIYGGKKFKVPEDDLNPNYVASITDAIYFAKP 460
Qy 118 YIDGRV-----YKVMDFLTKRLDITLSD-----QIRKIQKVNTYTPGKKIWL 161
Db 461 NMDKQLEEYNQVELLADLEPLAKITLSEMEETIGFTVDVHDLSEMEKEIOE-----KLDVLI 516
Qy 162 GVYVTSAGCTNNLSDSYAGFLWLTGLMLANOQIDVVIRHSFFDGHYNHLVD-----QN 216
Db 517 RNHDAAGDFNINPKQLGVVLFETLQ-----PVIKKT--KTGYSTAVDVLEQLQG 567
Qy 217 FNPDPYLSLLYKRLIGPKVLAVHAGLQKRPGRVIRDKLRIY 262
Db 568 EHPIDYILE--YROL--SKLQSTVVEGLQ-----KVISDDORIH 603

RESULT 14
S09811
hypothetical protein UL48 - human cytomegalovirus (strain AD169)
N:Alternate names: hypothetical protein HPRF0
C:Species: human cytomegalovirus, human herpesvirus 5
A:Note: host Homo sapiens (man)
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 08-Oct-1999
C:Accession: S09811
R:Chee, M.S.; Bankier, A.T.; Beck, S.; Bohni, R.; Brown, C.M.; Cerny, R.; Horsnell, T.; M.; Barrell, B.G.
Curr. Top. Microbiol. Immunol. 154, 125-169, 1990
A:Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus
A:Reference number: S09749; MUID:90269039
A:Accession: S09811
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2241 <CHE>
A:Cross-references: EMBL:X17403; NID:g95991; PIDN:CAA35407.1; PID:g1780826
A:Note: this sequence was submitted to the EMBL Data Library, December 1989

Query Match 4.7%; Score 92.5; DB 2; Length 2241;
Best Local Similarity 20.8%; Pred. No. 78;
Matches 78; Conservative 54; Mismatches 110; Indels 133; Gaps 21;

Qy 15 LQREKAQMHVLVLKQFSNTYSNLIITPE-----NNYRTMHGRAVNSQLGKDYIQ 66
Db 1325 LSAERTVRWMLAFLEALQ-----ITADTPHHSRTHYRNLLQQAVERAVTLAQHIE 1379
Qy 67 LKSLLOP-IRIYSRASLYGPNIGRPNKRNVIALLDGFMKVAGSVDAVTWQHCVIDGRV- 124
Db 1360 QNAACENFIAHQEATANGAS--TPRVDV-----QAVEAV-WORLE-PCRVA 1424
Qy 125 -----KVMDFLTKRLDITLSD-----QIRKIQKV 149
Db 1425 GAARHQKVOELLQ--RLGQTGLDELQETLATEFALLHGQTQTSYGLDFRSQLEKINDLR 1483
Qy 150 NTYTP-----GKKIWLQGV-----TTSAGCTNNLSDSYAGFLWLTGLMLANOQI 196
Db 1484 TRFAELAKRRGTRLSNKGVLNPRKPQATISLGA-----FTRGL 1522
Qy 197 DVVIRHSFFDGHY--NHLVDQNFNLPDYWLSLYKRLIGPKVL-AVHVAGLQKRPGR 253

Db 1523 NALERHVQLGHQYLLNKLNGS-----SLVYRLEDIPSVLPATH-----ETDPAL 1566
Qy 254 VIRDKLRIYAHCTNNHNNHNYVRGSIITLFIINLHRSRKKIKLAGTLRDKLVHQYLLQPYGQ 313
Db 1567 IMRDRLR--RLCFARHDTF-----LEVVDVFGMRQIVTQAG-----EPIH--LVTDYGN 1612
Qy 314 EGLKSKSVQLNGOPL 328
Db 1613 VAFKYLALRDDGRPL 1627

RESULT 15
S65187
GPI-anchor biosynthesis protein PIG-A - yeast (Saccharomyces cerevisiae)
N:Alternate names: GlcNAc-inositol phospholipid assembly protein; GPI-anchor biosynthesis
C:Species: Saccharomyces cerevisiae
C:Date: 10-Dec-1994 #sequence_revision 31-May-1996 #text_change 16-Jun-2000
C:Accession: S65187; S27131; A61112; S18842; S60357
R:Benes, V.; Rechmann, S.; Nentwich, U.; Voss, H.; Ansoerge, W.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S65183
A:Accession: S65187
A:Molecule type: DNA
A:Residues: 1-461 <BEN>
A:Cross-references: EMBL:Z73531; NID:g1370368; PIDN:CAA97882.1; PID:g1370369; GSPDB:G
A:Experimental source: strain S288C (AB972)
R:Fassler, J.
submitted to the EMBL Data Library, January 1992
A:Reference number: S27131
A:Accession: S27131
A:Molecule type: DNA
A:Residues: 16-461 <FAS1>
A:Cross-references: EMBL:X63290
A:Accession: A61112
A:Molecule type: DNA
A:Residues: 'MGF', 'II', 'IA', '16-461' <BAU>
A:Cross-references: EMBL:X63290; NID:g4533; PIDN:CAA44924.1; PID:g441475
R:Fassler, J.S.; Gray, W.; Lee, J.P.; Yu, G.; Gingerich, G.
Mol. Gen. Genet. 230, 310-320, 1991
A:Title: The Saccharomyces cerevisiae SPT14 gene is essential for normal expression o
A:Reference number: S18842; MUID:92079912
A:Accession: S18842
A:Molecule type: DNA
A:Residues: 16-429, 'S' <PAS2>
A:Cross-references: EMBL:X63290
R:Vossen, J.H.; Ram, A.F.J.; Klis, F.M.
Biochim. Biophys. Acta 1243, 549-551, 1995
A:Title: Identification of SPT14/CWH6 as the yeast homologue of hPIG-A, a gene involv
A:Reference number: S60357; MUID:95244613
A:Accession: S60357
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 16-461 <VOS>
C:Genetics:
A:Gene: SCD:SPT14; CWH6; GPI3; MIPS:YPL175W
A:Cross-references: SCD:S0006096; MIPS:YPL175W
A:Map position: 16L
A:Introns: 5/2
C:Superfamily: GPI-anchor biosynthesis protein PIG-A
C:Keywords: glycoprotein; transmembrane protein
F:416-434/Domain: transmembrane #status predicted <TMM>
F:140,386/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 4.7%; Score 92; DB 1; Length 461;
Best Local Similarity 22.0%; Pred. No. 9.2;
Matches 50; Conservative 38; Mismatches 63; Indels 76; Gaps 12;

Qy 181 GFLLWLTGLMLANOQIDVVIRHSFFDH---GYNHLVDQNFNLPDYWL----- 225
Db 117 GILHANTMG-----LRTVFTDHSLYGFNNLTSIWNKLLFTLTNIDRVICVST 166
Qy 226 ---SLLYKRLIGPKVLAV---HVAGLQKRP-----PGRVIRDKL-----RIYAH--- 264

```
Db 167 CKENHVRTELSPDIIISVIPNAVVSDFKPRDPTGGTKRKQSRDKIVIVIGRLFPNKG 226
Qy 265 -----CTNHHNHNY-VRGSITLFI-----INLHRSRKIKLACT-----LRDKL- 302
Db 227 DLLTRIIPKVCSSHEDVEFIVAGDGPKEFIDFQQMIESHRLQKRVQLLGSVPHEKVRDVL 286
Qy 303 -----VHQYLLQPYQEGGLKSKSVOLNGOPLVMVDDGTLPCLKPRPL 344
Db 287 QGDIYLHASLTFEFGTILVEAASCNL---LIVTTQVGGIPEVLPNEM 330
```

Search completed: July 30, 2002, 08:16:34
Job time: 351 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:31:33 ; Search time 36.14 Seconds
(without alignments)
399.623 Million cell updates/sec

Title: US-09-836-461-2_COPY_162_534

Perfect score: 1956

Sequence: 1 KGCKIAQHPDVMLEQLREKA.....PVTMGFFVKNYNALACRYR 373

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 segs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100.5	5.1	630	1	GIDA_TREPA
2	94	4.8	625	1	GIDA_LACIA
3	94	4.8	1517	1	RPOC_CAMJE
4	92.5	4.7	2241	1	TEGO_HCMVA
5	92	4.7	461	1	GP13_YEAST
6	92	4.7	534	1	COX1_KLULA
7	92	4.7	535	1	COX1_HANWI
8	91.5	4.7	335	1	SYW_BUCAT
9	90	4.6	1616	1	VITL_CABEL
10	88.5	4.5	286	1	PARB_CHLPN
11	88.5	4.5	341	1	YF43_MYCTU
12	88	4.5	1106	1	CIC2_RABIT
13	87.5	4.5	534	1	COX1_SACDO
14	87.5	4.5	557	1	COX1_NEUCR
15	87.5	4.5	627	1	MUTL_BAGSU
16	87.5	4.5	828	1	YFA4_YEAST
17	87	4.4	528	1	NIFK_FRAAL
18	87	4.4	535	1	GIDA_LACIC
19	85.5	4.4	512	1	COX1_YEAST
20	85.5	4.4	773	1	HEXB_ALTSO
21	85.5	4.4	898	1	ACOC_CUCMA
22	85	4.3	552	1	DP1B_ECOLI
23	84.5	4.3	434	1	TI47_HUMAN
24	84.5	4.3	928	1	CN2A_RAT
25	84	4.3	514	1	COX1_PROMI
26	84	4.3	621	1	GIDA_RICPR
27	84	4.3	905	1	COFG_SCHPO
28	83	4.2	401	1	ACHN_CHICK
29	83	4.2	536	1	DIT1_YEAST
30	83	4.2	838	1	SVV_MYCPN
31	83	4.2	934	1	SVL1_SULSO
32	82.5	4.2	527	1	COX1_OENBE
33	82.5	4.2	541	1	COX1_PODAN

34	82.5	4.2	564	1	DNL1_FOWPV	Q67480 fowlpox vir
35	82.5	4.2	921	1	CN2A_BOVIN	P14099 bos taurus
36	82.5	4.2	1065	1	SP18_SCHPO	O13710 schizosacch
37	82.5	4.2	1118	1	YIPL_YEAST	P40456 saccharomyc
38	82	4.2	527	1	COX1_SOYBN	P07506 glycine max
39	82	4.2	591	1	PAK4_HUMAN	O96013 homo sapien
40	82	4.2	854	1	AKA3_BOVIN	O77797 bos taurus
41	82	4.2	1399	1	RPOC_PSEAE	Q9HCW9 pseudomonas
42	82	4.2	1541	1	MRP2_RAT	Q63120 rattus norv
43	81.5	4.2	477	1	EXU2_DROPS	O24617 drosophila
44	81.5	4.2	645	1	LIP1_PHOLU	P40601 photorhabdu
45	81.5	4.2	687	1	TGLC_BOVIN	PS1176 bos taurus

ALIGNMENTS

RESULT 1						
GIDA_TREPA						
ID	GIDA_TREPA	STANDARD:	PRT:	630 AA.		
AC	O83084:					
DT	15-DEC-1998 (Rel. 37, Created)					
DT	15-DEC-1998 (Rel. 37, Last sequence update)					
DT	16-OCT-2001 (Rel. 40, Last annotation update)					
DE	Glucose inhibited division protein A.					
GN	GIDA OR TP0044.					
OS	Treponema pallidum.					
OC	Bacteria; Spirochaetales; Spirochaetaceae; Treponema.					
OX	NCBI_TaxID=160;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=NICHOLS;					
RX	MEDLINE=98332770; PubMed=9665876;					
RA	Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,					
RA	Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,					
RA	Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,					
RA	Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,					
RA	McDonald L., Artlach P., Bowman C., Cotton M.D., Fujii C., Garland S.,					
RA	Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,					
RA	Venter J.C.;					
RT	"Complete genome sequence of Treponema pallidum, the syphilis					
RT	spirochete.";					
RL	Science 281:375-388(1998).					
CC	-1- FUNCTION: NOT KNOWN.					
CC	-1- SIMILARITY: BELONGS TO THE GIDA FAMILY.					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/					
CC	or send an email to license@isb-sib.ch).					
CC	-----					
DR	EMBL; AE001189; AAC65038.1;					
DR	TIGR; TP0044;					
DR	InterPro; IPR002218; GIDA.					
DR	Pfam; PF01134; GIDA; 1.					
DR	ProDom; PD003738; GIDA; 1.					
DR	PROSITE; PS01280; GIDA_1; 1.					
DR	PROSITE; PS01281; GIDA_2; 1.					
KW	Complete proteome.					
SQ	SEQUENCE 630 AA; 70175 MW; 1B52C6F02C1EC275 CRC64;					

Query Match	5.1%;	Score 100.5;	DB 1;	Length 630;
Best Local Similarity	21.3%;	Pred. No. 1.1;		
Matches	77;	Conservative 52;	Mismatches 129;	Indels 103; Gaps 20;
Qy	57	GSOLCK--DYIOLKSLQPIRIYSRASLYGNIGRPKNVIALLDGFMKVAGSTVDVATW	114	
Db	68	GEMGKFKADACMIQ-----YRLNLSR--GPAVQAPRIQADKFL--YAKVKYITLECTQH	118	

Qy	115	QHCVIGRVVKVMDFLKTRLLDLSQIRKIOKVNVNTYTPGKKIMLEGVYVTTTSAGTNNL	174
Db	119	LHLQD-----TVDVVCSNTTDAGYVAYGAAHAVVT--ARGRRISARAVVLTTG-----	166
Qy	175	SDSYAGFLWL-----NTYGLMANOGIDVIRHSFFDHG-----YNHLVDQNF	217
Db	167	--TFMEGRVYIGEVAPEGRGLGEHAAGLAAALKKGFQMGRLKTGTPARVLRKSYD----	221
Qy	218	NPLPDYWLSLLYR-----LIGPKVLAVHVAGLORKPRPGRVIFDKLRIYAHCTNNHHNY	273
Db	222	-----LSYMEKQEADAIRPFESFA-HV-----EINRPHADCYINTNETHOL	263
Qy	274	VRGSITLFIINLHRS---RKIKLAGT-----LRDKLVHOYLLOPYGQEGLS	318
Db	264	IRE-----NFHRSPPFSGRKAVGTRYCPSTEDKVRFPDRIRHQLYIEP---EGLDT	313
Qy	319	KSVOLNGOPLVMVDD-----GTLPULK-----RPLRAGRTLVIPPVTMGF-FVVKNVNA	367
Db	314	EELINGLSSCLPEDIQEMIRTIPTGMRVITRPAYADVAVLFPVQLGIDILQTRKVS	373
Qy	368	L 368	
Db	374	L 374	

RESULT	2	GIDA_LACLA	STANDARD;	PRT;	625 AA.
ID		GIDA_LACLA			
AC		Q9CEJ4;			
DT	16-OCT-2001	(Rel. 40, Created)			
DT	16-OCT-2001	(Rel. 40, Last sequence update)			
DT	01-MAR-2002	(Rel. 41, Last annotation update)			
DE		Glucose inhibited division protein A.			
GN	GIDA OR LL1847.				
OS	Lactococcus lactis (subsp. lactis).				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;				
OC	Lactococcus.				
OX	NCBI_TaxID=1360;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=IL1403;				
RX	MEDLINE=21235186; PubMed=11337471;				
RA	Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K.,				
RA	Weissenbach J., Ehrlich S.D., Sorokin A.;				
RT	"The complete genome sequence of the lactic acid bacterium <i>Lactococcus</i>				
RT	<i>lactis</i> ssp. <i>lactis</i> IL1403."				
RL	Genome Res. 11:731-753(2001).				
CC	-1- FUNCTION: NOT KNOWN.				
CC	-1- SIMILARITY: BELONGS TO THE GIDA FAMILY.				

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC	EMBL; AE006414; AK05945.1; -; -;
DR	InterPro; IPR001327; FAD_pyr_redox.
DR	InterPro; IPR002218; GIDA.
DR	InterPro; IPR001100; pyr_redox.
DR	pfam; PF01114; GIDA; 1.
DR	PRINTS; P00368; FADPNR.
DR	PRINTS; P00411; PNDRTASEI.
DR	ProDom; PD003738; GIDA; 1.
DR	PROSITE; PS01280; GIDA_1; 1.
DR	PROSITE; PS01281; GIDA_2; 1.
KW	Complete proteome.
SO	SEQUENCE 625 AA; 69377 MW; D79C

Query Match 4.88; Score 94; DB 1; Length 625;

```

Best Local Similarity 19.68; Pred. No. 3.7;
Matches 72; Conservative 119; Indels 118; Gaps 17;

QY 68 KSLQPIRIYSRASL-YGPNIGRPKRKNVIAL-LDGFMKVAGSTVDATWQHICYID---GR 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 31 KTLMLTINLNMVAPWPCNPISGGSAKGIIVREIDALGALGEMGRNIDKTVIQMKMLATGKGP 90
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 123 VVKVM-----DFLKRLLDRLDSQ-----IRKI-----OKVNTYT-PGKKTLWL 160
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 91 AVRALRAQADKDEYADSMKNTVSDQENLTLRQGMVEELILDEEKKVKIGIKTSTGTKYGA 150
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 161 EGVVTT-----SAGGTNNLSDSYAAGFLMLNTLGLMLANQGIDVIVRHSF 204
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 151 KAVIITGTALRGEIIGELKYSSGPNNSL-----SIGLADN-----LREIG 193
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 205 FDGHYNHLVDQNFNLPDYWLSLLYKRLIGPKVLAVHVAGLQKPRGP-----252
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 194 FEIG-----RKTGTPPKVLASSIDYDKTEIQPDGEAPNHFSPNSS 234
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 253 ---RVIRDKLRIYAHCTNNHHNNVYRGSSITLFTIINLJHRS---RKKTKLAG-----TLRDKL 302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 235 DENYLKDQIPCWLTWTTTNSHTILRG-----NLHRAPLESGIVKGVGPRYCPSTIEDKI 287
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 303 V-----HQYLLOPYGOEGLKSKSVQLNGQBLVWVDDGTLPBLKPRPLRAGSTLVIPPV 355
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 288 TRFADKPRHQLFLEP---EGRNTEEVYIGGLSTSPEDVQVDFDLVKSIPGLENAQMRPGY 344
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 356 TMGFFVW 362
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 345 AIEYDVW 351
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 3
RPOC_CAMJE STANDARD; PRT; 1517 AA.
AC Q9PI30;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase
DE beta' chain) (RNA polymerase beta' subunit).
DE RPOC OR CJO479.
GN Campylobacter jejuni.
OS Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OC NCBI_TaxID=197;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=NTC 11168;
RC MEDLINE=20150912; PubMed=10688204;
RX Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -!- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
CC BETA' CHAIN.
CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```


RX MEDLINE=95255221; PubMed=7737116;
 RA Schoenbacher M., Horvath A., Fassler J.S., Riezman H.;
 RT "The yeast spt14 gene is homologous to the human PIG-A gene and is
 RL required for GPI anchor synthesis.";
 CC EMBO J. 14:1637-1645(1995).
 CC -1- FUNCTION: NECESSARY FOR THE SYNTHESIS OF N-ACETYLGALACTOSAMINYL-
 CC BIOSYNTHESIS.
 CC -1- PHOSPHATIDYLINOSITOL, THE VERY EARLY INTERMEDIATE IN GPI-ANCHOR
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.
 CC -1- CAUTION: WAS ORIGINALLY (REF.1) THOUGHT TO BE INVOLVED IN
 CC TRANSCRIPTION.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; X63290; CAA44924.1; -;
 DR EMBL; Z73531; CAA97882.1; -;
 DR PIR; S27131; S27131.
 DR PIR; S18842; S18842.
 DR SGD; S0006096; SPT14.
 DR InterPro; IPR001296; Glycos_transf_1.
 DR Pfam; PF00534; Glycos_transf_1; 1.
 KW Transferase; Glycosyltransferase.
 FT CONFLICT 1 15 MSSSKHVESYNENR -> MGFNIA (IN REF. 1).
 SQ SEQUENCE 461 AA; 52372 MW; 3DDC44DD5DA43E5A CRC64;

Query Match 4.7%; Score 92; DB 1; Length 461;

Best Local Similarity 22.08; Pred. No. 3.6;
Matches 50; Conservative 38; Mismatches 63; Indels 76; Gaps 12;

Qy 181 GFLWNTLGLMANGIDVIRHSFDH---GYNHLVDGNFNPDPYL----- 225
 Db 117 GILHANTMG-----LRTVETDLSLGFNLTSTIWNKLLTFTLTNIDRVICVSN 166
 Qy 226 ---SLLYKRLGPKVLAV---HVAGLQRPKPR-----PGRVIRDKL----- 264
 Db 167 KENMIVTELSPPDIISIPNAVSEDPKRDPTGGTKRKOSDKIVIVIGRFPNKG 226
 Qy 265 -----CTNNHNNHY-VRGSTITLFI-----INLHRSRKKIKLAGT-----LRDKL- 302
 Db 227 DLLTRIIPKVCSSHEDVEFVAGDGPKEFDQMTESHRLQKRVQLLGSVPHEKVRDVL 286
 Qy 303 -----VHQYLLOPQGEGLKSKSVOLNGQPLVMVDDGTLPKLPPL 344
 Db 287 QGDYILHASLAEATGILVEAASCNL---LIVTTQVGGIPEVLPNEM 330

RESULT 6

COX1_KLJULA
 ID COX1_KLJULA STANDARD; PRT; 534 AA.
 AC P20386;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
 GN COX1.
 OS Kluyveromyces lactis (Yeast).
 OC Mitochondrion.
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
 OX NCBI_Taxid=28985;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K8;
 RX MEDLINE=92035081; PubMed=1657415;
 RA Hardy C.M., Clark-Walker G.D.;
 RT "Nucleotide sequence of the COX1 gene in Kluyveromyces lactis

RT mitochondrial DNA: evidence for recent horizontal transfer of a group
 RT II intron.";
 RL Curr. Genet. 20:99-114(1991).
 RN [2]
 RC SEQUENCE OF 1-29 FROM N.A.
 RP STRAIN=K8;
 RX MEDLINE=91021490; PubMed=2171241;
 RA Hardy C.M., Clark-Walker G.D.;
 RT "Nucleotide sequence of the cytochrome oxidase subunit 2 and val-crna
 RT genes and surrounding sequences from Kluyveromyces lactis K8
 RT mitochondrial DNA.";
 CC yeast 6:403-410(1990).
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
 CC AND COPPER B.
 CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferrocyclochrome
 CC c + 2 H(2)O.
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; X57546; CAA40765.1; -;
 DR EMBL; X15999; CAA34130.1; -;
 DR HSPSP; P00396; IOCC.
 DR InterPro; IPR000883; COX1.
 DR Pfam; PF00115; COX1; 1.
 DR PRINTS; PR01165; CYCOXIDASE1.
 DR PROSITE; PS00077; COX1; 1.
 DR Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
 KW Respiratory chain; Inner membrane.
 KW METAL 62 62 IRON (HEME A) (PROBABLE).
 FT METAL 241 241 COPPER B (PROBABLE).
 FT METAL 245 245 COPPER B (PROBABLE).
 FT METAL 290 290 COPPER B (PROBABLE).
 FT METAL 291 291 COPPER B (PROBABLE).
 FT METAL 376 376 IRON (HEME A3) (PROBABLE).
 FT METAL 378 378 IRON (HEME A) (PROBABLE).
 SQ SEQUENCE 534 AA; 58677 MW; 4A26B1C323F9F1C2 CRC64;

Query Match 4.7%; Score 92; DB 1; Length 534;

Best Local Similarity 22.98; Pred. No. 4.4;
Matches 72; Conservative 39; Mismatches 97; Indels 106; Gaps 19;

Qy 74 IRIYSRASLYG-----PNIGRPKNVIALLDGFMKVGAGSTVDATVW-QHCYIDGRV 124
 Db 259 VSTYSKRPVFGESVMYAMASIG-----LLGFL-----VNSHMIYVG--- 296
 Qy 125 KVMDFLTKRLDITLSDQIRKTKVNTYTPGKKIWLGEVVTTSAGGTNNLSDS---YAAGF 182
 Db 297 --LD-ADTRAVFTSATMIIAIPGTGIKIFS-----WL-----ATIVGGSLRAVPMYATAF 344
 Qy 183 LWLNTLG-----MLANQGDVIRHSFDHGYNHLY---DQNFNPLPDYW-----LSLL 228
 Db 345 LFLFTIGLGTGVALANASLDVAFDHTDYYVVGHFHYLSGMAIFSLFAGYVNSPOILGLY 404
 Qy 229 YKR-----LIGPKV--LAVHVGALQRPKPRGRIROKLRVIAHCTNNHNNHYVRG 276
 Db 405 YNEKLAQTQFWLIFGANVIFLPMHFLGVNGMPR-----RIPDPDAFAGNHYVAS 455

```

Qy  277 -----STLFTIINLRSKKIKLAGTRDKLVHGYLL-----QPVGQBLGK 317
      :| | | :       :| | | | :| | :| :
Db   456 IGSIIAIVSFLEFLYILD-----QLVNGLENKNRKSVIYNKGPDVFESNQIFATNKIK 510
      :| | | :       :| | | :| | :| :
Qy   318 SKSQV--LNGQPLV 329
      :| | :| | | |
Db    511 SSSIEFLTSPPAV 524
      :| | :| | | |

RESULT          7
COX1_HANWI
ID COX1_HANWI STANDARD; PRT; 535 AA.
AC P48868;
DT 01-FEB-1996 (Rel. 33, Created)
DD 01-FEB-1996 (Rel. 33, Last sequence update)
DE 01-FEB-1996 (Rel. 33, Last annotation update)
DN Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
GN COXI
OS Hansenula winegei (Yeast).
OG Mitochondrion.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
ON NCBI_TaxID=4907;
RX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=21;
RA Sekito T., Okamoto K., Kitano H., Yoshida K.;
RC "Yeast Hansenula winegei mitochondria genome's complete DNA sequence
RD demonstrated unique characteristics."
RE Nucleic Acids Symp. Ser. 31:233-234(1994).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B.
CC -1- CATALYTIC ACTIVITY: 4 ferrocytochrome c + O(2) = 4 ferricytochrome
CC c + 2 H(2)O.
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to licensel@isb-sib.ch).
-----
DR EMBL; D31785; BAA06563.1; .
DR HSPP; P00396; LOCC.
DR InterPro; IPR000883; COX1.
DR Pfam; PF00115; COXI; 1.
DR PRINTS; PR01165; CYCOXIDASEI.
DR PROSITE; PS00077; COXI; 1.
KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW Respiratory chain; Inner membrane.
FT METAL 63 63 IRON (HEME A) (PROBABLE).
FT METAL 242 242 COPPER B (PROBABLE).
FT METAL 246 246 COPPER B (PROBABLE).
FT METAL 291 291 COPPER B (PROBABLE).
FT METAL 292 292 COPPER B (PROBABLE).
FT METAL 377 377 IRON (HEME A3) (PROBABLE).
FT METAL 379 379 IRON (HEME A) (PROBABLE).
SQ SEQUENCE 535 AA; 59416 MW; 5302237917443D38 CRC64;

Query Match          4.78; Score 92; DB 1; Length 535;
Best Local Similarity 22.7%; Pred. No. 4.4;
Matches 69; Conservative 42; Mismatches 103; Indels 90; Gaps 19;
```

Query Match 4.7%; Score 92; DB 1; Length 535;
Best Local Similarity 22.7%; Pred. No. 4.4;
Matches 69; Conservative 42; Mismatches 103; Indels 90; Gaps 19;

Qy	74	IRIYSRASLYGPNIGRPNKYNIALLDGFMKVGSTVDATVM-QHCYIDGRVVKVYMDFLKT	132
Db	260	VSTYSKRPVFGEMSMVYAMASIAFL-GFL-----VMSHHMIVG-----LD-ADT	302
Qy	133	RLDITLSDIQRKIKQVNTYTPGKKIMLGVVTTTSGAGTNNLS--DSYAAGFLWLNTLG-	189
Db	303	RAYTSSTMVJAVPTGIMKFS-----WL-----ATLYGGSIRLAVPMTYATIAFLFTMGG	353
Qy	190	-----MLANQGIDVIRHSFFDHGYNHLV---DQNFNPLPDYW-----LSLLYKR-----	231
Db	354	LTGVALANASLDVAFHDYVYVVGHHFVLSMGAIFSLFAGYVYVWSPQILGLYFNERLAQM	413
Qy	232	-----LIGPKV--LAVHHVAGLQRPRGRVTRDKLRIYAHCTNNHHNNHYVRG-----	276
Db	414	QFWLIFGVAGVIFMPMHPLGTQGMPR-----RMPDYPDAYAGWNYVSSMGSVMAIM	464
Qy	277	SITFIINLHRSKKIKLAGTLRDLKLVHQ---YLLQP-----YGOEGLKSKSVO--L	323
Db	465	SLAFLIYIMYD-----QLINGLTNKMNDKSVVYSKAPDFVESNTWFAANSIKSASIEFL	519
Qy	324	NGQP 327	
Db	520	NSPP 523	
RESULT 8			
SYW_BUCAI			
ID	SYW_BUCAI	STANDARD; PRT; 335 AA.	
AC	P57602;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Tryptophan--tRNA ligase) (TPRS).		
DE	TPRS OR P57602.		
GN	Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum symbiotic bacterium).		
OS	Bacteria; Proteobacteria; gamma subdivision; Buchnera.		
OC	NCBI_TaxID=118099;		
OX	[1]		
RN	SEQUENCE FROM N.A.		
RP	STRAIN-TOKYO 1998;		
RC	MEDLINE=20445173; PubMed=10993077;		
RX	Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;		
RA	"Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS.";		
RT	Nature 407:81-86(2000).		
RL	-!- CATALYTIC ACTIVITY: ATP + L-tryptophan + tRNA(Trp) = AMP + dihydrophosphate + L-tryptophanyl-tRNA(Trp).		
CC	-!- SUBUNIT: HOMODIMER (BY SIMILARITY).		
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic.		
CC	-!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-tRNA SYNTHETASE FAMILY.		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; AP001119; BAB13229.1; -		
DR	InterPro; IPR001412; trna-synt_I.		
DR	InterPro; IPR002306; trna-synt_trp.		
DR	Pfam; PF00579; trna-synt_ib; 1.		
DR	PRINTS; PR01039; TRNASYNTHTRP.		
DR	PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.		
KW	Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding; Complete proteome.		
FT	SITE 14 22		"HIGH" REGION.
FT	SITE 196 200		"KMSKS" REGION.
FT	BINDING 199 199		ATP (BY SIMILARITY).

KW	Complete proteome.		
FT	FT SITE	14	"HIGH" REGION.
FT	FT SITE	196	"KMSKS" REGION.
FT	FT BINDING	199	ATP (BY SIMILARITY).

```
SQ SEQUENCE 335 AA; 38549 MW; 871562D5A6734E3F CRC64;

Query Match
Best Local Similarity 4.7%; Score 91.5; DB 1; Length 335;
Matches 46; Conservative 33; Mismatches 64; Indels 65; Gaps 10;

QY 22 OMHLVL---LKEQFSNTYSNLLITENNRYRTMHGRAVNSQLGKDYIOLKSLQPIRIYS 78
DB 152 KQVELTRIAIRFNSLYGH-VFTLPKPLITQHS-----KINSLLSPSKMS 198
QY 79 RASLYGPNIGRPRKNVIALLDG---FMKVAGSTVDVATWQHCYID-----GRVVKVM 127
DB 199 KSDI-----NKNVIFLLDDIKTVISKIQAAYTSETPSKIYYDIEKKPGISNLLLEIL 251
QY 128 -----DFL-----KRLDTLSDQIRKIOKVNTYTPG-----KKWLEGV 163
DB 252 SAITNKDDILLKEGLMYSEFNKIVADHLSKFLYKLOKSYRNDENVYLLKTIAYEGA 311
QY 164 VTTSGGTTNNLSDSYAAGFLWNTLQML 191
DB 312 MKSQLSKNTLTKVY-----DKLGLI 332

RESULT 9
VIT1_CABEL STANDARD; PRT; 1616 AA.
AC P55155;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Vitellogenin 1 precursor.
GN VIT-1 OR K09F5.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Bentley D.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PRECURSOR OF THE EGG-YOLK PROTEINS THAT ARE SOURCES OF
CC NUTRIENTS DURING EMBRYONIC DEVELOPMENT (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U37430; AAB52675.1; -.
DR WormPep; K09F5.2; CE04746.
DR InterPro; IPR001747; Vitellogenin_N.
DR InterPro; IPR001846; Vwd.
DR Pfam; PF01347; Vitellogenin_N; 1.
DR Pfam; PF00094; vwd; 1.
DR SMART; SM00216; VWD; 1.
KW Storage protein; Multigene family; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1616 VITELLOGENIN 1.
SQ SEQUENCE 1616 AA; 188063 MW; F173D7C452D123F5 CRC64;

Query Match
Best Local Similarity 4.6%; Score 90; DB 1; Length 1616;
Matches 76; Conservative 68; Mismatches 131; Indels 96; Gaps 17;

QY 26 VLLKEQFSNTY---SNLLITPENRYRTMHGRAVNSQLGKDYIOLKSLQPIRIYSRASV 83
DB 233 VLIRPQTVTYIILENELKE-SEVRSLYTVNVNGQEVMTKTRSKLVLE--ENHSIKS-H 288
```

```
QY 84 GPNIGRPRKNVI-----ALLDGEKMKVAGSTVDVATWQHCYIDGRVVKVMDFLKTRLLDT 137
DB 289 LEKVNGEKESIIYSRWELQVDEDFK-NGDKAEFAFPFEKFLD-----KKMHLTKT 338
QY 138 LSDQIRKIQKVNTYTPGKKIWLEGV---TTSAGGTNNLSDSYAAGFLWNTLGMLAN 193
DB 339 ITEIQIEVENNI---PETSFLARLVRIFRTS---TSQKEIH-----ETLYVKAD 384
QY 194 QGIDVVIRHSF-----FDHGYNHLVDONENPLDPYWLSLLYKRLIGPKVLAVHVAGL 245
DB 385 KKIQSLMEHALAIAIACTKNTIQHILVHIENEDIVPLE-----AAQLKSI 428
QY 246 QRKPPGRVIRDKRIYAHCTNHHNNHYVR-----GSITLFIINLHRSR----- 289
DB 429 QETPFSQTIARALIKFAESRVSKNNQVRSQAWLAAGSVVRGIVDYKNIRPLVREDKRE 488
QY 290 -----KKIKIAGTLRDKLVHOYLLQPYQGGLSKSVQNLNGQPLVWVDDTTLPEL 339
DB 489 LKEKELRVFMQYKDAETTYEKIL---ALKSIGNAGLDTISVNLNE---IIVDKRQLLPV 542
QY 340 KPRPLRAGRTL 350
DB 543 RKEAIDALRLL 553

RESULT 10
PARB_CHLPN STANDARD; PRT; 286 AA.
AC Q9Z7M0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable chromosome partitioning protein parB.
GN PARB OR CPN0684 OR CP0062.
OS Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamyidophila.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CWL029;
RX MEDLINE-99206606; PubMed-10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AR39;
RX MEDLINE-20150255; PubMed-10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heldelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-J138;
RX MEDLINE-20330349; PubMed-10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -!- FUNCTION: INVOLVED IN CHROMOSOME PARTITION. LOCALIZE TO BOTH POLES
CC OF THE PREDIVISIONAL CELL FOLLOWING COMPLETION OF DNA REPLICATION.
CC BINDS TO THE DNA ORIGIN OF REPLICATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE PARB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```


OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88336904; PubMed=2458626;
 RA Ellis S.B., Williams M.E., Ways N.R., Brenner R., Sharp A.H.,
 RA Leung A.T., Campbell K.P., McKenna E., Koch W.J., Hui A.,
 RA Schwartz A., Harpold M.M.;
 RT "Sequence and expression of mRNAs encoding the alpha 1 and alpha 2
 RT subunits of a DHP-sensitive calcium channel.";
 RL Science 241:1661-1664(1988).
 RN [2]
 RP SEQUENCE OF 961-973.
 RX MEDLINE=91131638; PubMed=1847144;
 RA Jay S.D., Sharp A.H., Kahl S.D., Vedvick T.S., Harpold M.M.,
 RA Campbell K.P.;
 RT "Structural characterization of the dihydropyridine-sensitive calcium
 RT channel alpha 2-subunit and the associated delta peptides.";
 RL J. Biol. Chem. 266:3287-3293(1991).
 [3]
 RN SEQUENCE OF 961-975; 992-1000 AND 1033-1050.
 RX MEDLINE=90368635; PubMed=2168391;
 RA de Jongh K.S., Warner C., Catterall W.A.;
 RA "Subunits of purified calcium channels. Alpha 2 and delta are encoded
 RT by the same gene.";
 RL J. Biol. Chem. 265:14738-14741(1990).
 CC -!- FUNCTION: CALCIUM CHANNEL PROTEIN WHICH PLAYS AN IMPORTANT ROLE IN
 CC EXCITATION-CONTRACTION COUPLING.
 CC -!- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FOUR SUBUNITS:
 CC ALPHA-1, ALPHA-2, BETA AND GAMMA. ALPHA-2 AND DELTA FORMS
 CC HETERODIMERS THAT ARE DISULFIDE-LINKED.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS ARE PRODUCED BY ALTERNATIVE
 CC SPLICING.
 CC -!- TISSUE SPECIFICITY: SKELETAL MUSCLE.
 CC -!- PTM: ALPHA-2 AND DELTA SUBUNITS ARE PROTEOLYTICALLY PROCESSED FROM
 CC A PRECURSOR FORM.
 CC -!- SIMILARITY: TO OTHERS SPECIES ALPHA-2 SUBUNIT.
 CC -!- SIMILARITY: CONTAINS 1 VFMA DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M21948; AAA81562.1; -;
 DR PIR; S10579; CHRA2.
 DR InterPro; IPR004010; Cache.
 DR InterPro; IPR002035; VFMA.
 DR Pfam; PF02743; Cache; 1.
 DR Pfam; PF00092; vwa; 1.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS50234; VFMA; 1.
 KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
 KW Calcium channel; Glycoprotein; Phosphorylation; Signal;
 KW Alternative splicing.
 FT SIGNAL 1 26
 FT CHAIN 27 960
 FT CHAIN 961 1106
 FT TRANSMEM 448 471
 FT TRANSMEM 921 945
 FT TRANSMEM 1082 1101
 FT DOMAIN 255 432
 FT CARBOHYD 94 94
 FT CARBOHYD 138 138
 FT CARBOHYD 186 186
 FT CARBOHYD 326 326
 FT CARBOHYD 350 350
 FT CARBOHYD 477 477
 FT CARBOHYD 606 606

FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 678 678 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 784 784 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 827 827 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 851 851 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 898 898 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 988 988 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1001 1001 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT MOD_RES 503 503 PHOSPHORYLATION (BY CAPK)
 FT MOD_RES 848 848 PHOSPHORYLATION (BY CAPK)
 FT MOD_RES 1106 1106 AA; 125042 MW; BOODE7F3C877B618 CRC64;
 FT SQ SEQUENCE 1106 AA; 125042 MW; BOODE7F3C877B618 CRC64;
 Query Match 4.5%; Score 88; DB 1; Length 1106;
 Best Local Similarity 18.7%; Pred. No. 24; Indels 122; Gaps 14;
 Matches 68; Conservative 55; Mismatches 119; Indels 122; Gaps 14;
 QY 92 KNVIALDGFPMKVGSTVDVAVTWQHCYIDGRVVKVMDFLKRLDLTLD----- 140
 Db 254 KDLMLIVD-----VSGS-----VSGTLKLRISVSSEMLETSDDDDFNVASFNS 298
 QY 141 ---QIRKIQKVVNTYTPGKIMLEGVVTTSAGGTNNLSDSYAAGFLWL-----NT 187
 Db 299 NAQDVSCFQHLVQANVRNKKVLKDAVNNITAKGITYKKGFSFAFQOLLNINVSRANCK 358
 QY 188 LGMLANOGIDVIRHSHFDHGYNHLV-----DONFNPDPDWLSL----- 227
 Db 359 IIMFTDGGERAQEIFAKYNKKDKVRFVSVQHNVDGRPIOMACENKGYVVEIPSI 418
 QY 228 -----LYKRLIG-PKVL-----VHVAGLQKRPGRVIRDKLRIYACTN 267
 Db 419 GAIRINTQEVLDVGRPMVLADGAKAQVQWTVNYLDLLEL-----GLVITGTLFVNTIGQ 474
 QY 268 HHNNYVRSITLFIINLHRSRKKIKLAGTLRDLKVLQYLPYQGEGLSKSVQLANGQP 327
 Db 475 FENTNLKNQILGVMGVDVSLEDIK-----RLTPFTLCPNGY---YFAIDPNQYV 523
 QY 328 LVMVDDGTLPKLPRLRAGRTLVP-----PVTMGFFVFNVNVALA 369
 Db 524 LLH-----PNLQPKPIGVG----IPTINLKRPNVQNPQKSBPVTLDLFDLDAELENDIK 573
 QY 370 CRYR 373
 Db 574 VEIR 577
 RESULT 13
 COXI_SACDO
 ID COXI_SACDO STANDARD; PRT; 534 AA.
 AC P98001;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
 GN COXI.
 OS Saccharomyces douglasii (Yeast).
 OG Mitochondrion.
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=46617;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SD12;
 RX MEDLINE=93185919; PubMed=8383070;
 RA Tian G.L., Michel F., Macadre C., Lazowska J.;
 RT "Sequence of the mitochondrial gene encoding subunit I of cytochrome
 RT oxidase in Saccharomyces douglasii.";
 RL Gene 124:153-163(1993).
 CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE

FT METAL 383 383 IRON (HEME A) (PROBABLE).
FT LIPID 324 324 MYRISTATE.
SQ SEQUENCE 557 AA; 61494 MW; D4D9618704E5DF6C CRC64;

Query Match 4.5%; Score 87.5; DB 1; Length 557;
Best Local Similarity 22.0%; Pred. No. 11;
Matches 46; Conservative 33; Mismatches 69; Indels 61; Gaps 13;
QY 74 IRIYSRASLYGNIGRPNKRVNIALDGFEMKVGASTVDVW-OHCYIDGRVVKVMDFLKT 132
DB 264 ISAYSNKSVFG-----YIGWVYAMSI--GILGEIWSHHMYTVG-----LD-VDT 306
QY 133 RLIDTLDSDQIRKIQKVVNTYTPGKKIWLGVVTTGAGTNNLSDS--YAAGFLWLNTLG- 189
DB 307 RAYFTAATLIIAVPTGKIFS-----WL-----ATCYGGSIRLTPSMLFALGFVFMFTIGG 357
QY 190 ----MLANOGIDVIRHSFP-----DHGYNHLVDO-----NFNPL---P 221
DB 358 LSGVVLNANSLDIAFDHTIYVVAHFHYVLSMGAVFAMFSGWYHWPKILGLNAYNVLSKA 417
QY 222 DWLSLLYKRLGPKVLAVHVHAGLQKRP 250
DB 418 QEWLLFIGNL---TFPQHFLGLQGMPR 443

RESULT 15
MUTL_BACSU STANDARD; PRT; 627 AA.
ID AC P49850;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA mismatch repair protein mutL.
GN MUTL.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_Taxid=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96349107; PubMed=8760914;
RA Ginetti F., Perego M., Albertini A.M., Galizzi A.
RT "Bacillus subtilis mutL operon: identification, nucleotide sequence and mutagenesis."
RL Microbiology 142:2021-2029(1996).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE REPAIR OF MISMATCHES IN DNA. IT IS REQUIRED FOR DAM-DEPENDENT METHYL-DIRECTED DNA MISMATCH REPAIR. MAY ACT AS A "MOLECULAR MATCHMAKER". A PROTEIN THAT PROMOTES THE FORMATION OF A STABLE COMPLEX BETWEEN TWO OR MORE DNA-BINDING PROTEINS IN AN ATP-DEPENDENT MANNER WITHOUT ITSELF BEING PART OF A FINAL EFFECTOR COMPLEX (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTL/HEXB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U27343; AAB19236.1; -.
DR EMBL; Z99112; CAB13578.1; -.
DR HSSP; P23367; 1BKN.
DR Subtilist; BG11402; mutL.
DR InterPro; IPR002099; DNA_mis_repair.
DR InterPro; IPR003594; HATPase_C.
DR InterPro; IPR004359; HIS_KIN_sig.
DR Pfam; PF01119; DNA_mis_repair; 1.
DR Pfam; PF02518; HATPase_c; 1.
DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.

KW DNA repair; Complete proteome.
SQ SEQUENCE 627 AA; 70431 MW; 068A0509CC265343 CRC64;

Query Match 4.5%; Score 87.5; DB 1; Length 627;
Best Local Similarity 22.8%; Pred. No. 13;
Matches 70; Conservative 43; Mismatches 107; Indels 87; Gaps 17;
QY 128 DFLKRLDITLSDQIRKTKQV-----VNTYT---PGKKIWLGE--VV-----TTSAGGT--- 171
DB 86 DLFVRVTLGFRGEALPSIASVSHLEITTTSTGEGAGTKLVLOGGNIISRSRSSRKGTEIV 145
QY 172 -NNLSDSYAAGFLWNT-----LG-----MLANOGIDVIRHSFHDGYNHLVDQNF 217
DB 146 VSNLEFNTPARLKYMKYVHTELGNITDVVNRIALAHPEVSIIRLH-----HGKN-LLOQNG 200
QY 218 NPLPDYWLSSLYKRLIGPKVLAVHVAGLQ-----RKPR-----PGRVIR 257
DB 201 NGDVRHVLAAIYGTAVALKMLPLHVSSLDPEVKYIALPEITRASRNMYSSVYNGRYIKN 260
QY 258 KLRIYAHCTNHH-----NHNVVRGSIIT---LFIINLHRSRKKIKLA-----GTLRD- 300
DB 261 PFLVRAVHEGYHTLLPIGRHPITFTIEYTMDPILDVNVNHPSKLEVLRSKETELHDLIRDG 320
QY 301 --KLVRHQLLPYGOEGLKSKSVQLNQPLVMVDD-----GTLPELKPRLRAGR 348
DB 321 IKDVFQOOLIPSAQVPKSAFAKNEQQQFITFDEKPEKKVPEKSTAPSYSPMKL---S 377
QY 349 TLVIPPV 355
DB 378 SVVKEPV 384

Search completed: July 30, 2002, 08:31:35
Job time: 982 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:30:49 ; Search time 123.94 Seconds
(without alignments)
520.632 Million cell updates/sec

Title: US-09-836-461-2-copy_162_534
Perfect score: 1956
Sequence: 1 KGCKTAQHPDVMLEQREKA.....PVTMGFFVKNYNALACRYR 373

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPREMBL_19.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mmc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1956	100.0	534	4 Q9HB38	Q9hb38 homo sapien
2	1917	98.0	592	4 Q9HB37	Q9hb37 homo sapien
3	1743	89.1	480	4 Q9HB39	Q9hb39 homo sapien
4	810.5	41.4	536	11 Q9QZF8	Q9qzf8 rattus norv
5	787.5	40.3	543	4 Q9Y251	Q9y251 homo sapien
6	787.5	40.3	545	4 Q9UL39	Q9ul39 homo sapien
7	778	39.8	545	6 Q9MYV0	Q9myv0 bos taurus
8	736.5	37.7	523	13 Q90YK5	Q90yk5 gallus gall
9	301	15.4	521	10 Q9SDA1	Q9sda1 arabisdopsis
10	301	15.4	543	10 Q9FF10	Q9ff10 arabisdopsis
11	258.5	13.2	516	10 Q9FLK8	Q9flk8 arabisdopsis
12	236	12.1	536	10 Q9FZP1	Q9fzp1 arabisdopsis
13	215.5	11.0	527	10 Q9LRC8	Q9lrc8 scutellaria
14	155	7.9	190	10 Q82604	Q82604 arabisdopsis
15	155	7.9	935	5 Q9VE79	Q9ve79 drosophila
16	110	5.6	174	10 Q9ATW5	Q9atw5 zea mays (m

ALIGNMENTS

RESULT 1

ID	Q9HB38	PRELIMINARY;	PRT;	534 AA.
AC	Q9HB38;			
DT	01-MAR-2001 (TREMBlrel. 16, Created)			
DT	01-MAR-2001 (TREMBlrel. 16, Last sequence update)			
DT	01-MAR-2001 (TREMBlrel. 16, Last annotation update)			
DE	HEPARANASE-LIKE PROTEIN HPA2B.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20483645; PubMed=11027606;			
RA	McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,			
RA	Hircock M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;			
RT	"Cloning and Expression Profiling of Hpa2, a Novel Mammalian			
RT	Heparanase Family Member."			
RL	Biochem. Biophys. Res. Commun. 276:1170-1177(2000).			
DR	EMBL; AF282886; AAG23422.1; .			
SQ	SEQUENCE 534 AA; 60063 MW; C3DE5E900CB338C4 CRC64;			

Query Match	Best Local Similarity	Matches	373;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	KGCKTAQHPDVMLEQREKAQMHLVLLKQFSNTYSNLILTEPNRYTMHGRAVNSQL	60								
Db	162	KGCKTAQHPDVMLEQREKAQMHLVLLKQFSNTYSNLILTEPNRYTMHGRAVNSQL	221								
Qy	61	KGKDIQLKSLLOPIRIYSRASLYGNIGPRKNVIALLDGFMKVGSTVDATWQHCYID	120								
Db	222	KGKDIQLKSLLOPIRIYSRASLYGNIGPRKNVIALLDGFMKVGSTVDATWQHCYID	281								
Qy	121	GRVVKVMDFLKRLDLSQIRKIQKVVNTYTPGKIKWLEGVTTSGGTTNLSDSVAA	180								
Db	282	GRVVKVMDFLKRLDLSQIRKIQKVVNTYTPGKIKWLEGVTTSGGTTNLSDSVAA	341								

QY 181 GFLWNTLGMLANQIDVIRHSFFDHGYNHLVDQNFNPLDPYWLKRLIGPKVLAV 240
Db 342 GFLWNTLGMLANQIDVIRHSFFDHGYNHLVDQNFNPLDPYWLKRLIGPKVLAV 401
QY 241 HVAGLQRPGRVIRDKLRIYAHCTNHHNNHNVVRSITLFIINLHRSRKKIKLAGTLRD 300
Db 402 HVAGLQRPGRVIRDKLRIYAHCTNHHNNHNVVRSITLFIINLHRSRKKIKLAGTLRD 461
QY 301 KLVIHQYLQPYGQGLKSKSVQNLGQPLVMVDDGTLPELKPRLRAGRTLVIPTVMGFF 360
Db 462 KLVIHQYLQPYGQGLKSKSVQNLGQPLVMVDDGTLPELKPRLRAGRTLVIPTVMGFF 521
QY 361 VVKVNNALACRYR 373
Db 522 VVKVNNALACRYR 534
RESULT 2
Q9HB37 PRELIMINARY; PRT; 592 AA.
AC Q9HB37:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE HEPARANASE-LIKE PROTEIN HPA2C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20483645; PubMed=11027606;
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,
RA Hircok M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;
RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian
RT Heparanase Family Member."
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).
DR EMBL: AF282887; AAG23423.1; -.
SQ SEQUENCE 592 AA; 66580 MW; 95C384AD9A74258E CRC64;

Query Match 98.0%; Score 1917; DB 4; Length 592;
Best Local Similarity 86.5%; Pred. No. 1.1e-158;
Matches 373; Conservative 0; Mismatches 0; Indels 58; Gaps 1;
QY 1 KGCKIAQHPDVMLELQREAAQHLLVLLKEQFSNTYSNLLIT----- 42
Db 162 KGCKIAQHPDVMLELQREAAQHLLVLLKEQFSNTYSNLLITARSLDKLYNFADCSGLHL 221
QY 43 -----EPNNYRTMHGRAVNGSQLGK 62
Db 222 IFALNLRNPNNSWSSSALSLLKYSASKYINISHELGNPNYRTMHGRAVNGSQLGK 281
QY 63 DYIQLKSLQPIRIYRASLYGNIGRPRKNVIALDGMKVGSTVDVAVTQHCYIDGR 122
Db 282 DYIQLKSLQPIRIYRASLYGNIGRPRKNVIALDGMKVGSTVDVAVTQHCYIDGR 341
QY 123 VVKVMDFLKRLDLTSDQIRKTKQKVNTYTPGKKIWLGVVTTISAGGTNNLSDSYAAGF 182
Db 342 VVKVMDFLKRLDLTSDQIRKTKQKVNTYTPGKKIWLGVVTTISAGGTNNLSDSYAAGF 401
QY 183 LWNLTGMLANQIDVIRHSFFDHGYNHLVDQNFNPLDPYWLKRLIGPKVLAVHV 242
Db 402 LWNLTGMLANQIDVIRHSFFDHGYNHLVDQNFNPLDPYWLKRLIGPKVLAVHV 461
QY 243 AGLQRPGRVIRDKLRIYAHCTNHHNNHNVVRSITLFIINLHRSRKKIKLAGTLRDKL 302
Db 462 AGLQRPGRVIRDKLRIYAHCTNHHNNHNVVRSITLFIINLHRSRKKIKLAGTLRDKL 521
QY 303 VHQYLLQPYGQGLKSKSVQNLGQPLVMVDDGTLPELKPRLRAGRTLVIPTVMGFFV 362
Db 522 VHQYLLQPYGQGLKSKSVQNLGQPLVMVDDGTLPELKPRLRAGRTLVIPTVMGFFV 581

QY 363 KVNVALACRYR 373
Db 582 KVNVALACRYR 592
RESULT 3
Q9HB39 PRELIMINARY; PRT; 480 AA.
AC Q9HB39:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE HEPARANASE-LIKE PROTEIN HPA2A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20483645; PubMed=11027606;
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,
RA Hircok M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;
RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian
RT Heparanase Family Member."
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).
DR EMBL: AF282885; AAG23421.1; -.
SQ SEQUENCE 480 AA; 53900 MW; F75F89F67AC1FF83 CRC64;
Query Match 89.1%; Score 1743; DB 4; Length 480;
Best Local Similarity 100.0%; Pred. No. 1.2e-143;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 43 EPNNYRTMHGRAVNGSQLGKDYIQLKSLQPIRIYRASLYGNIGRPRKNVIALDGMF 102
Db 150 EPNNYRTMHGRAVNGSQLGKDYIQLKSLQPIRIYRASLYGNIGRPRKNVIALDGMF 209
QY 103 KVAGSTVDVAVTQHCYIDGRVVKVMDFLKRLDLTSDQIRKTKQKVNTYTPGKKIWL 162
Db 210 KVAGSTVDVAVTQHCYIDGRVVKVMDFLKRLDLTSDQIRKTKQKVNTYTPGKKIWL 269
QY 163 VVTSAGTNNLSDSYAAGFLWNTLGMLANQIDVIRHSFFDHGYNHLVDQNFNPLD 222
Db 270 VVTSAGTNNLSDSYAAGFLWNTLGMLANQIDVIRHSFFDHGYNHLVDQNFNPLD 329
QY 223 YWLSLLYKRLIGPKVLAVHVAGLQRPGRVIRDKLRIYAHCTNHHNNHNVVRSITLFI 282
Db 330 YWLSLLYKRLIGPKVLAVHVAGLQRPGRVIRDKLRIYAHCTNHHNNHNVVRSITLFI 389
QY 283 INLHRSRKKIKLAGTLRDKLVIHQYLQPYGQGLKSKSVQNLGQPLVMVDDGTLPELKP 342
Db 390 INLHRSRKKIKLAGTLRDKLVIHQYLQPYGQGLKSKSVQNLGQPLVMVDDGTLPELKP 449
QY 343 PLRAGRTLVIPTVMGFFVVKVNNALACRYR 373
Db 450 PLRAGRTLVIPTVMGFFVVKVNNALACRYR 480
RESULT 4
Q9QZF8 PRELIMINARY; PRT; 536 AA.
AC Q9QZF8:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HEPARANASE.
OS HEP.
GN Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]

```
RP SEQUENCE FROM N.A.
RA Podyma K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;
RT "Heparanase from parathyroid cell line.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF184967; AAF04563.1; -.
SQ SEQUENCE 536 AA; 60569 MW; 6208B1FD9EE28421 CRC64;

Query Match 41.4%; Score 810.5; DB 11; Length 536;
Best Local Similarity 40.6%; Pred. No. 3.5e-62;
Matches 170; Conservative 66; Mismatches 114; Indels 69; Gaps 6;

QY 10 DYMLELQREKAAQMHVLLKKEQF-----SNTYS-----NLIL- 41
DB 128 DVLRLQMEWPFQ-ELLLRLQYQREFKNSTYSRSVDMLYSFAKCSRLDLIFGLNALLR 186
QY 42 -----TEPNRYTMHGRAVNGSOLGKDYIQLKSL 71
DB 187 TPDLRWSSNAQLLLNYCSSKGYNISWELGNEPNSFWKKAISIDGLQLGDFVELHKL 246
QY 72 QPIRYSRASLYGNIGRPNKRVIALLDGFMKVGSTVDATVWQHCYIDGRVVKVYDFLK 131
DB 247 QK-SAFQNAKLYGPDIGOPRGKTKVLLRSFLKAGGEVIDSLTWHYVYLNGRVATKEDFLS 305
QY 132 TRLLDTLSQDTRKTKQVNTVTPGKKIWLEGVVTTSAGTNNLSDSYAAGFLWLNTLQML 191
DB 306 SDVLDTFILSVQKILKVTKENTPGKKVWLGETSSAYGGGAPLLSNTFAAGFMWLDKLGLS 365
QY 192 ANQIDVYIRHSFFDHGYNHLVDQNFNPLPDYWLSSLKRLIGPKVLAVHAGLQKRP 251
DB 366 AOLGIEVVRQVFGAGNYHLVDENFELPDYWLSSLKRLIGPKVLAVHAGLQKRP 420
QY 252 GRVIRDKLRIYAHCTNHHNNHYVRSITLFIINLHRSRKKIKLAGTLRDKLVHQLQPY 311
DB 421 ----RSKLRYVLLHCTNVYHPRYREGDLYLVNLHNVTKLKLPPMFSPRPVKYLLKPF 476
QY 312 GOEGLKSKSVQLNGOPLVWVDGTLPELKPRLRAGRTLVPVPTMGFFVVKVNNALAC 370
DB 477 GSDGLLSKSVQLNGOPLVWVDGTLPELKPRLRAGRTLVPVPTMGFFVVKVNNALAC 535

RESULT 5
QY9251 PRELIMINARY; PRT; 543 AA.
AC QY9251;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE HEPARANASE.
GN HPA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=99321249; PubMed=10395326;
RA Hulet M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,
RA Parish C.R.;
RT "Cloning of mammalian heparanase, an important enzyme in tumor
RT invasion and metastasis.";
RL Nat. Med. 5:803-809 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Vlodavsky I., Friedman Y., Eskin M., Aingorn H., Atznorn R.,
RA Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I.,
RA Spector L., Pecker I.;
RT "Mammalian heparanase: a novel gene involved in tumor progression and
RT metastasis.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
```

```
RX MEDLINE=99377052; PubMed=10446189;
RA Toyoshima M., Nakajima M.;
RT "Human heparanase. Purification, characterization, cloning, and
RT expression.";
RL J. Biol. Chem. 274:24153-24160 (1999).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=99333379; PubMed=10405343;
RA Kussie P.H., Hulmes J.D., Ludwig D., Patel S., Navarro E.C.,
RA Seddon A.P., Giorgio N.A., Bohlen P.;
RT "Cloning and Functional Expression of a Human Heparanase Gene.";
RL Biochem. Biophys. Res. Commun. 261:183-187 (1999).
DR EMBL; AF165154; AAD45379.1; -.
DR EMBL; AF144325; AAD41342.1; -.
DR EMBL; AF155510; AAD54941.1; -.
DR EMBL; AF152376; AAD45669.1; -.
SQ SEQUENCE 543 AA; 61176 MW; AD262EC267334AB2 CRC64;

Query Match 40.3%; Score 787.5; DB 4; Length 543;
Best Local Similarity 43.9%; Pred. No. 3.6e-60;
Matches 156; Conservative 59; Mismatches 115; Indels 25; Gaps 3;

QY 31 QFSNTYSNLIL-----TEPNRYTMHGRAVNGSOLGKDYIQLKSLQPIR 75
DB 198 QWNSNAQLLLDYCSSKGYNISWELGNEPNSFLKKADIFINGSOLGEDFIQLKLLRK-S 256
QY 76 IYSRASLYGNIGRPNKRVIALLDGFMKVGSTVDATVWQHCYIDGRVVKVWDFLKTRLL 135
DB 257 TPKNAKLYGPDVGPQPRKTKVLLRSFLKAGGEVIDSVTHIHYVYLNGRVATREDFLNPDL 316
QY 136 DTLSQDTRKTKQVNTVTPGKKIWLEGVVTTSAGTNNLSDSYAAGFLWLNTLQMLANOG 195
DB 317 DIFISSVQKVFQVSTPRGKKVWLGETSSAYGGGAPLLSNTFAAGFMWLDKLGLSARMG 376
QY 196 IDVTRHSFFDHGYNHLVDQNFNPLPDYWLSSLKRLIGPKVLAVHAGLQKRPGRVI 255
DB 377 IEVVRQVFGAGNYHLVDENFELPDYWLSSLKRLIGPKVLAVHAGLQKRPGRVI 429
QY 256 RDKLRIYAHCTNHHNNHYVRSITLFIINLHRSRKKIKLAGTLRDKLVHQLQPYQGBG 315
DB 430 --KLRYVLLHCTNTPRYKEGDTLYLVNLHNVTKLKLPPMFSPRPVKYLLRPLGPHG 487
QY 316 LKSKSVQLNGOPLVWVDGTLPELKPRLRAGRTLVPVPTMGFFVVKVNNALAC 370
DB 488 LLSKSVQLNGOPLVWVDGTLPELKPRLRAGRTLVPVPTMGFFVVKVNNALAC 542

RESULT 6
QY9UL39 PRELIMINARY; PRT; 545 AA.
AC QY9UL39;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HEPARANASE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=20229546; PubMed=10764835;
RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;
RT "Heparanase expression in invasive trophoblasts and acute vascular
RT damage.";
RL Glycobiology 10:467-475 (2000).
DR EMBL; AF084467; AAD54516.1; -.
SQ SEQUENCE 545 AA; 61418 MW; 67B80ACD73C5A9A1 CRC64;
```


AC O9SDAL:
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE HYPOTHETICAL 57.8 KDA PROTEIN.
GN F13G24.30.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Van Der Schueren J., Chuang Y.J., Voet M., Robben J.,
RA Volckaert G., Bancroft I., Meves H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL133421; CAB62595.1; -.
DR InterPro; IPR001254; Trypsin.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 521 AA; 57831 MW; 07D8664A4B305CC2 CRC64;

Query Match 15.4%; Score 301; DB 10; Length 521;
Best Local Similarity 28.2%; Pred. No. 9.8e-18;
Matches 105; Conservative 50; Mismatches 127; Indels 90; Gaps 15;

QY 52 GRAVNGSQLGKDYIOLKSLQPIRIYSRASLYGPNIGRPRKNVIALLDGF-----M 102
Db 186 GASVSAELYGKDLIVLKDVIN--KVYKNSWLHKPILVAP-----GGFYEQQWYTKLL 235
QY 103 KVAG-STVDVAVTQHCYIDGR-----VVKVMDFLKTRLLDLSOIRKIQKVNTYTPG 155
Db 236 EISGPSVVDVVT-HHIYNLGSNDPALVKKIMD---PSYLSQVSKTFKDVNQTIQEHGP- 290
QY 156 KKIWLEGVVVTSAG---GTNNLSDSYAAGFLWNLTLGMLANOGIDVIRHSEFFDHGYNH 211
Db 291 ---WASPMWGESGGAYNSGGRHVSTFIDSFYLDQLGMSARHNTKVYCRQLVGGFYGL 347
QY 212 LVDQNFNPLPDYWLKLLKRLGPKVLAVHVAGLQKRPGRVIRDKLRIYAHCTNHHNH 271
Db 348 LEKGFVFPNPDYISALLWHLRMKGVLAVQTDG-----PPQLRVYAHCSKG--- 393
QY 272 NVYRGSIITLFIINLH-----RSRKKIKLAGTLRDLKLV----- 303
Db 394 ---RAGVTLLLLINLSNQSDFTVSVSNGINVLNAESRKKKSLDRTLKRPFSWIGSKASDG 450
QY 304 ----HQYLLQPYGQEG-LKSKSVQLNGQPLVMVDDGTLPKLPRLRAGRTLVIPVPTWG 358
Db 451 YLNREYHLP--ENGVLRSKTMVLNGSKLKTATGDIPLSLEPVLRSVNSPLNVLPLSMS 508
QY 359 FFWKNNVNALAC 370
Db 509 FIVLPNFDASAC 520

RESULT 10
Q9FF10 PRELIMINARY; PRT; 543 AA.
AC Q9FF10;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SIMILARITY TO HEPARANASE.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=97471969; PubMed=9330910;
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
RT features of the 1.6 Mb regions covered by twenty physically assigned
RT pl clones";
RL DNA Res. 4:215-230(1997).
DR EMBL; AB005249; BAB09947.1; -.
DR InterPro; IPR001254; Trypsin.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
SQ SEQUENCE 543 AA; 60250 MW; 0FA2248948282FF6 CRC64;

Query Match 15.4%; Score 301; DB 10; Length 543;
Best Local Similarity 28.2%; Pred. No. 1e-17;
Matches 105; Conservative 50; Mismatches 127; Indels 90; Gaps 15;

QY 52 GRAVNGSQLGKDYIOLKSLQPIRIYSRASLYGPNIGRPRKNVIALLDGF-----M 102
Db 208 GASVSAELYGKDLIVLKDVIN--KVYKNSWLHKPILVAP-----GGFYEQQWYTKLL 257
QY 103 KVAG-STVDVAVTQHCYIDGR-----VVKVMDFLKTRLLDLSOIRKIQKVNTYTPG 155
Db 258 EISGPSVVDVVT-HHIYNLGSNDPALVKKIMD---PSYLSQVSKTFKDVNQTIQEHGP- 312
QY 156 KKIWLEGVVVTSAG---GTNNLSDSYAAGFLWNLTLGMLANOGIDVIRHSEFFDHGYNH 211
Db 313 ---WASPMWGESGGAYNSGGRHVSTFIDSFYLDQLGMSARHNTKVYCRQLVGGFYGL 369
QY 212 LVDQNFNPLPDYWLKLLKRLGPKVLAVHVAGLQKRPGRVIRDKLRIYAHCTNHHNH 271
Db 370 LEKGFVFPNPDYISALLWHLRMKGVLAVQTDG-----PPQLRVYAHCSKG--- 415
QY 272 NVYRGSIITLFIINLH-----RSRKKIKLAGTLRDLKLV----- 303
Db 416 ---RAGVTLLLLINLSNQSDFTVSVSNGINVLNAESRKKKSLDRTLKRPFSWIGSKASDG 472
QY 304 ----HQYLLQPYGQEG-LKSKSVQLNGQPLVMVDDGTLPKLPRLRAGRTLVIPVPTWG 358
Db 473 YLNREYHLP--ENGVLRSKTMVLNGSKLKTATGDIPLSLEPVLRSVNSPLNVLPLSMS 530
QY 359 FFWKNNVNALAC 370
Db 531 FIVLPNFDASAC 542

RESULT 11
Q9FLK8 PRELIMINARY; PRT; 516 AA.
AC Q9FLK8;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SIMILARITY TO HEPARANASE.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=98290546; PubMed=9628582;
RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
RT Sequence features of the regions of 1,456,315 bp covered by nineteen
RT physically assigned pl and TAC clones";
RL DNA Res. 5:41-54(1998).
DR EMBL; AB010073; BAB08480.1; -.

DR InterPro: IPR001917; AminoTransf_2.
DR PROSITE: PS00599; AA_TRANSFER_CLASS_2; UNKNOWN_1.
SQ SEQUENCE 516 AA; 57414 MW; 6943E6D3D89FB93C CRC64;

Query Match 13.2%; Score 258.5; DB 10; Length 516;
Best Local Similarity 23.7%; Pred. No. 5e-14;
Matches 94; Conservative 63; Mismatches 130; Indels 109; Gaps 14;

QY 46 NYRTMHGRAVN-----GSQL-----GKDYIOLKSLQPIRIYSRASLYGPN 87
DB 158 NYTVSGYALDSWFGNELSGGSIWASVVELYKGLIVLKNVKNYKNSRT----- 210
QY 88 GRPRKNVIALLDG-----MKVAGSTVDVAVTWHQCHYIDG-----RVKVMDFLKT 132
DB 211 ----KPLVAPAGFFEEQWYSELLRLSGPCVLDVLTTHIYNLGPNGDPKLVNKILD---P 263
QY 133 RLDTDSQIRKTKQVNVNTYTPGKKIWLGVVTSAGTNN-----LSDSYAAGFWLWNT 187
DB 264 NYLSGISSELFANVQTIQEHGPWAAAV-----GEAGGAFNSGGROVSETFINSEFWYLDQ 318
QY 188 LGMLANQGDVIRHSFFDHGYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHVAGLQR 247
DB 319 LGISSKHNTKVCYCRQALVGFGYGLLEKETFPVNPDPYSALLWHLRMKGILGVQTTA--- 375
QY 248 KPRPGRVIRDKLRIYAHCTNHHNHNVRGSIITLFIINLHR-----S 288
DB 376 -----SEYLRAYVHCKSR-----RAGITILLNLKSKHTFTVAVSGVKVVLQAES 421
QY 289 RKIKIAGTLRDKLV-----HQYLLQPYGQEG-LKSKSVQLNGQPLVMVDDG 334
DB 422 MKRKSFLKTIKSVWGVNGKADGYLNREYHLSP--KGDLSKIMLLNGKPLVPTATG 479
QY 335 TLPELAPRPLRAGRTLVIPTVMGFFVKNVNALAC 370
DB 480 DIPKLEPVRHGVKSPVYINPLSISFVLPTFDAPAC 515

RESULT 12
Q9FZP1 PRELIMINARY; PRT; 536 AA.
AC Q9FZP1; (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SIMILARITY TO HEPARANASE
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB028613; BAB10787.1; -.
SQ SEQUENCE 536 AA; 59654 MW; 24202B8E82F3DB0E CRC64;

Query Match 12.1%; Score 236; DB 10; Length 536;
Best Local Similarity 22.8%; Pred. No. 4.8e-12;
Matches 91; Conservative 66; Mismatches 141; Indels 102; Gaps 16;
QY 32 FSNTYSNLIITPNYR-----TMHGRAYNSQLGKDYIQLKSLLPRIYSR 79
DB 177 YTNAESFIRTAENNTYIDGWELGSGVGARVANOYAIDTINLRNVN--RVYKN 234
QY 80 AS-----LYGPNIGRPRKNVIALLDGPKV-----AGSTVDVAVTWHQCHY-----IDG 121
DB 235 VSPMPLVIGPG-----GFEVDWFTYLNKAENSLNATT-RHIYDLGPGVDE 280

QY 122 RVKVMDFLKTRLDLSQIRKIQKVWNTYTPGKKIWLGVVVTSG-----GTNNLSDS 177
DB 281 HLIE--KILNPSYLDQEAQKFRSLKNIKNSSTKAVAW-----VGESGGAYNSGRNLVSNA 334
QY 178 YAAGFWLWNTLGMLANQGDVIRHSFFDHGYNHLVDQNFNPLPDYWLSSLYKRLIGPKV 237
DB 335 FVYSEFWYLDQGLGSMASLYDKTCYRQSLIGGLTNTFTNPDPYSALINRQLMGRKA 394
QY 238 LAVHVAGLQKRPGRVIRDKLRIYAHCTNHHNHNVRGSIITLFIINLHR---KKIKL 294
DB 395 LFTTFSGTK-----KIRSYTHCARQSK-----GITVLLMLNDONTTIVVAKVEL 437
QY 295 AGTLRLDKLVHQYLLQPY-----GQBG-----LKSKSVQLNGQPLVM 330
DB 438 NNSF--SLRHTKHKYSKRASSQLFGPGNGVQIOREEVHLTAKDGNLHLSQTMLLNGALQV 495
QY 331 VDDGTLPELKPRLRAGRTLVIPTVMGFFVKNVNALAC 370
DB 496 NSMGDLPPITPIETHINSTEPTITTIAPYSIVFVHMNRNVVPAC 535

RESULT 13
Q9LRC8 PRELIMINARY; PRT; 527 AA.
AC Q9LRC8;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE BETA-GLUCURONIDASE.
GN SGUS.
OS Scutellaria baicalensis.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Lamiales; Lamiaceae; Scutellaria.
OX NCBI_TaxID=65409;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20418130; PubMed=10858442;
RA Sasaki K., Taura F., Shoyama Y., Morimoto S.;
RT "Molecular Characterization of a Novel beta-Glucuronidase from
Scutellaria baicalensis Georgi.";
RL J. Biol. Chem. 275:27466-27472(2000).
DR EMBL; AB040072; BAA97804.1; .
DR InterPro: IPR001179; FKBP_PP1ase.
DR PROSITE: PS00453; FKBP_PP1ASE_1; UNKNOWN_1.
SQ SEQUENCE 527 AA; 58772 MW; A5DE7C423F2A1E2B CRC64;

Query Match 11.0%; Score 215.5; DB 10; Length 527;
Best Local Similarity 28.5%; Pred. No. 2.9e-10;
Matches 70; Conservative 38; Mismatches 99; Indels 39; Gaps 10;
QY 128 DFLKTRLLD-TLSDQIRK-----IQKVVNTYTPGKK--TWLEGVVTTSAGTNNLSDSYA 179
DB 289 DALKDVLTLTASFDEATKSMYELQKIVN--RPGTKAVAWIGBAGGAFNSGQDGISNTFI 346
QY 180 AGLFWLNTLGMLANQGDVIRHSFFDHGYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLA 239
DB 347 NGFWYLNMLGYSALLDTKTCRQTLTGNGYGLLQTTIYPNPDPYSALLWHRUMGSKVLK 406
QY 240 VHVAGLQKRPGRVIRDKLRIYAHCTNHHNHNVRGSIITLFIINLHRSKKIKLA---- 295
DB 407 TEIVGTGK-----NVYIYAHCAKSN-----GITMLVLN-HDGESSVKISLDPS 448
QY 296 --GTLRDKLVHQYLLQPYGOEGLKSKSVQLNGQPLVMVDDGTLPELKPRLRAGRTLVI 353
DB 449 KYGSKRE----EYHLTPV--NNNLQSLRLVKLNGELLHLDPSGVIPALNPVEKDNKOLEVA 503
QY 354 PYTMGF 359
DB 504 PYSFMF 509

.Qy 361 VVKNNALACRY 372
|: :|| |::
Db 384 VLPDVNLEHCQF 395

Search completed: July 30, 2002, 08:30:51
Job time: 1003 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:15:07 ; Search time 53.59 Seconds
(without alignments)
170.008 Million cell updates/sec

Title: US-09-836-461-2_COPY_162_534

Perfect score: 1956

Sequence: 1 KGCKIAQHPDVMLQLEKA.....PVTMGFFVVKYNALACRYR 373

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents_AA:*
- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
 - 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
 - 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
 - 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
 - 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
 - 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	805	41.2	380	4	US-09-181-336-19
2	791.5	40.5	543	2	US-08-922-1708-10
3	791.5	40.5	543	4	US-09-071-7398-2
4	791.5	40.5	543	4	US-09-260-038B-2
5	787.5	40.3	543	4	US-09-181-336-13
6	779	39.8	380	4	US-09-181-336-17
7	744.5	38.1	532	4	US-09-181-336-15
8	88	4.5	1106	1	US-08-435-675B-5
9	88	4.5	1106	1	US-08-336-257A-8
10	86	4.4	1103	1	US-08-455-543A-53
11	86	4.4	1103	2	US-08-223-305C-53
12	84.5	4.3	434	2	US-08-989-925-1
13	84	4.3	538	4	US-09-175-928-4
14	82.5	4.2	921	1	US-07-872-644-39
15	82.5	4.2	921	1	US-08-297-494-39
16	82.5	4.2	921	1	US-08-297-510-39
17	82.5	4.2	921	1	US-08-479-532-39
18	82.5	4.2	921	1	US-08-455-526-39
19	82.5	4.2	921	1	US-08-455-525-39
20	82.5	4.2	921	3	US-09-139-491-39
21	82.5	4.2	921	5	PCT-US92-03222-39
22	82.5	4.2	942	1	US-07-872-644-43
23	82.5	4.2	942	1	US-08-297-494-43
24	82.5	4.2	942	1	US-08-297-510-43
25	82.5	4.2	942	1	US-08-479-532-43
26	82.5	4.2	942	1	US-08-455-526-43
27	82.5	4.2	942	1	US-08-455-525-43

28	82.5	4.2	942	3	US-09-139-491-43	Sequence 43, Appl
29	82.5	4.2	942	5	PCT-US92-03222-43	Sequence 43, Appl
30	82	4.2	591	3	US-09-082-737-2	Sequence 2, Appl
31	81	4.1	685	2	US-08-878-989-1	Sequence 1, Appl
32	81	4.1	685	3	US-09-136-282-2	Sequence 2, Appl
33	81	4.1	685	4	US-09-272-796-1	Sequence 1, Appl
34	81	4.1	685	4	US-09-505-744-2	Sequence 2, Appl
35	81	4.1	1788	2	US-08-962-284-2	Sequence 2, Appl
36	80.5	4.1	1086	6	5386025-8	Patent No. 5386025
37	80	4.1	313	1	US-08-302-449-2	Sequence 2, Appl
38	80	4.1	313	5	PCT-US94-07430-2	Sequence 2, Appl
39	80	4.1	1093	5	PCT-US93-03077-1	Sequence 1, Appl
40	79	4.0	934	1	US-08-215-805A-80	Sequence 80, Appl
41	79	4.0	1541	4	US-08-296-791-3	Sequence 3, Appl
42	79	4.0	1541	5	PCT-US95-10661A-3	Sequence 3, Appl
43	78.5	4.0	313	1	US-08-302-449-4	Sequence 4, Appl
44	78.5	4.0	313	5	PCT-US94-07430-4	Sequence 4, Appl
45	78.5	4.0	687	5	PCT-US91-09784-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-181-336-19
; Sequence 19, Application US/09181336A
; Patent No. 6242238
; GENERAL INFORMATION:
; APPLICANT: FREEMAN, Craig Geoffrey
; APPLICANT: PARISH, Richard
; APPLICANT: HUMORE, Brenton James
; APPLICANT: HULETT, Mark Darren
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN
; TITLE OF INVENTION: ENDOGLUCURONIDASE AND USES THEREFOR
; FILE REFERENCE: 032505-004
; CURRENT APPLICATION NUMBER: US/09/181.336A
; CURRENT FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: AU PP0062
; EARLIER FILING DATE: 1997-10-28
; EARLIER APPLICATION NUMBER: AU PP0812
; EARLIER FILING DATE: 1997-12-09
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 19
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Rattus sp.
US-09-181-336-19

Query Match	41.2%	Score 805;	DB 4;	Length 380;
Best Local Similarity	47.3%	Pred No. 5.1e-80;		
Matches 155;	Conservative 57;	Mismatches 106;	Indels 10;	Gaps 2;
QY	43	EPNNYRTMHGRAVNSQLGKDYIQLKSLIQLPIRISYRSASLYGNIPGRKKNVIALLDGPM	102	
Db	62	EPNSFWKKAHISIDGLQGLGDFVFLHKLQK-SAQNAKLYGPDIGQPRGKTVKLLRSFL	120	
QY	103	KVAGSTVDATVWQHGYIDGRVYKVMDFLKLRLDLSDOIRKIQKVVNTYTPGKKIWLGE	162	
Db	121	KAGGVEIDSLTWHHYLYLNGRVATKEDFLSSDVLDTFILSVQKILKVTKMTGCKVWLGE	180	
QY	163	VYTTSGAGTNNLSDSYAAGFLMLNTLGLMNOGIDVIRHSFFDGHYNHLDQNFNPLPD	222	
Db	181	TSSAYGGGAPLSDTFAAGFMWDLGLSLAQGLIEVVMRQVFFGAGNYHLVDENEPLPD	240	
QY	223	YWLSLLYKRLIGPKVLAVHAGLQRPGRVIRDKRLIYAHCTNHHNNHYVRSITLFI	282	
Db	241	YWLSLLFKKLVGPKVLMRSVKGPD-----RSKRLVYLHCTNVYHPRYREGDITLV	291	
QY	283	INLHRSRKIKUAGLIRDKLVHQLYLLQPYQGQGLSKSVQLNGQPLVWVDDGTLPELKPR	342	
Db	292	LNLHNVTKHLKPPPMFSPVDKYLKLPFGSDCLLSKSVQLNGQTLKMWDEQTLPALTEK	351	

Db 317 DIFISSVQKVFQVVESTRPGKRVWLGTSAYGGGAPLLSDTFAAGFWMKDLGLSARMG 376
Qy 196 IDVYIRHSFFDHGYNHLVDQNFNPLPDYWLSSLLYKRLGPKVLAVHVAGLQKRPGRVI 255
Db 377 IEVVMQVFFGAGNHLVDENFDPLPDYWLSSLLYKRLGPKVLAVHVAGLQKRPGRVI 429
Qy 256 RDKLRIYAHCTNHHNHNHNVVRSITLFIINLHRSRKKIKLAGTLDKLVHQLQPYGQEG 315
Db 430 --KLRYLHCTNTPRYKEGDLTLIYAINLHNVTYKLRPLPYFSSNKQVDKYLRLPLGPHG 487
Qy 316 LKSKSVQNLGLOPLVMVDDGTLPKLPRLAGRTLVIPVMTGFFVKNVNALAC 370
Db 488 LLSKSVQNLGTLKMWDDQTLPLMEKPLPGSSGLGPAFVSFFVIRNAKVAAC 542

RESULT 4
US-09-260-038B-2
; Sequence 2, Application US/09260038B
; Patent No. 6348344
; GENERAL INFORMATION:
; APPLICANT: Maty Ayal-Hershkovitz et al.
; TITLE OF INVENTION: GENETICALLY MODIFIED CELLS AND METHODS FOR
; EXPRESSING RECOMBINANT HEPARANASE
; AND METHODS OF PURIFYING SAME
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
; STREET: 2001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted to
; an ASCII file
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/260,038B
; FILING DATE: 02-Mar-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/071,618
; FILING DATE: May 1, 1998
; APPLICATION NUMBER: 09/071,739
; FILING DATE: May 1, 1998
; APPLICATION NUMBER: 08/922,180
; FILING DATE: September 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedman, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 910/16
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 972-3-562553
; TELEFAX: 972-3-562554
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-260-038B-2

Query Match 40.5%; Score 791.5; DB 4; Length 543;
Best Local Similarity 44.2%; Pred. No. 2.7e-78;
Matches 157; Conservative 58; Mismatches 115; Indels 25; Gaps 3;

Qy 31 QFSNTYSNLIL-----TEPNYRTMHGRAVNGSQLGKDYIQLKSLQPIR 75
Db 198 QWNSSNAQLLLDYCSSKGYNISWELGNPNFSLKKADIFINGSQLGEDYIQLHKLRLK-S 256
Qy 76 IYSRASLYGPNIGRPNKVNIALLDGFMKVAGSTVDATVWQHICYIDGRVVYKVMDFLKTLL 135
Db 257 TEKNAKLYGPDVGQPPRTAKMLKSFLLKAGGEVIDSVTHHHYLLNGRTATREDFLNPDL 316
Qy 136 DPLSDQIRKIQKVVNTYTPGKKIWLGVVTTTSGAGTNNLSDSYAAGFLWNLTLGLMANOG 195
Db 317 DIFISSVQKVFQVVESTRPGKRVWLGTSAYGGGAPLLSDTFAAGFWMKDLGLSARMG 376
Qy 196 IDVYIRHSFFDHGYNHLVDQNFNPLPDYWLSSLLYKRLGPKVLAVHVAGLQKRPGRVI 255
Db 377 IEVVMQVFFGAGNHLVDENFDPLPDYWLSSLLYKRLGPKVLAVHVAGLQKRPGRVI 429
Qy 256 RDKLRIYAHCTNHHNHNHNVVRSITLFIINLHRSRKKIKLAGTLDKLVHQLQPYGQEG 315
Db 430 --KLRYLHCTNTPRYKEGDLTLIYAINLHNVTYKLRPLPYFSSNKQVDKYLRLPLGPHG 487
Qy 316 LKSKSVQNLGLOPLVMVDDGTLPKLPRLAGRTLVIPVMTGFFVKNVNALAC 370
Db 488 LLSKSVQNLGTLKMWDDQTLPLMEKPLPGSSGLGPAFVSFFVIRNAKVAAC 542

RESULT 5
US-09-181-336-13
; Sequence 13, Application US/09181336A
; Patent No. 6242238
; GENERAL INFORMATION:
; APPLICANT: FREEMAN, Craig Geoffrey
; APPLICANT: PARISH, Richard
; APPLICANT: HAMDORF, Brenton James
; APPLICANT: HULETT, Mark Darren
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN
; TITLE OF INVENTION: ENDOGLUCURONIDASE AND USES THEREFOR
; FILE REFERENCE: 032505-004
; CURRENT APPLICATION NUMBER: US/09/181,336A
; CURRENT FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: AU PP0062
; EARLIER FILING DATE: 1997-10-28
; EARLIER APPLICATION NUMBER: AU PP0812
; EARLIER FILING DATE: 1997-12-09
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-181-336-13

Query Match 40.3%; Score 787.5; DB 4; Length 543;
Best Local Similarity 43.9%; Pred. No. 7.6e-78;
Matches 156; Conservative 59; Mismatches 115; Indels 25; Gaps 3;

Qy 31 QFSNTYSNLIL-----TEPNYRTMHGRAVNGSQLGKDYIQLKSLQPIR 75
Db 198 QWNSSNAQLLLDYCSSKGYNISWELGNPNFSLKKADIFINGSQLGEDYIQLHKLRLK-S 256
Qy 76 IYSRASLYGPNIGRPNKVNIALLDGFMKVAGSTVDATVWQHICYIDGRVVYKVMDFLKTLL 135
Db 257 TEKNAKLYGPDVGQPPRTAKMLKSFLLKAGGEVIDSVTHHHYLLNGRTATREDFLNPDL 316
Qy 136 DPLSDQIRKIQKVVNTYTPGKKIWLGVVTTTSGAGTNNLSDSYAAGFLWNLTLGLMANOG 195
Db 317 DIFISSVQKVFQVVESTRPGKRVWLGTSAYGGGAPLLSDTFAAGFWMKDLGLSARMG 376
Qy 196 IDVYIRHSFFDHGYNHLVDQNFNPLPDYWLSSLLYKRLGPKVLAVHVAGLQKRPGRVI 255
Db 377 IEVVMQVFFGAGNHLVDENFDPLPDYWLSSLLYKRLGPKVLAVHVAGLQKRPGRVI 429
Qy 256 RDKLRIYAHCTNHHNHNHNVVRSITLFIINLHRSRKKIKLAGTLDKLVHQLQPYGQEG 315

Db 430 --KLRYLHCTNTDPRYKEGDLTVYALNHNVTYKRLPYFSPKQVDKYLRLPLGPHG 487
QY 316 LKSKSVQLNGQPLVMVDGTLPELKPRLRAGRTLVIPTVMGFVFFVKNVNALAC 370
Db 488 LLSKSVQLNGLTLMKVDQTLPLMEKPLRPGSSGLGPAFYSFFVIIRNAKYAAC 542

RESULT 6

US-09-181-336-17
; Sequence 17, Application US/09181336A
; Patent No. 6242238
; GENERAL INFORMATION:
; APPLICANT: FREEMAN, Craig Geoffrey
; APPLICANT: PARISH, Richard
; APPLICANT: HAMDOFF, Brenton James
; APPLICANT: HULETT, Mark Darren
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN
; FILE REFERENCE: 032505-004
; CURRENT APPLICATION NUMBER: US/09/181,336A
; CURRENT FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: AU PP0062
; EARLIER FILING DATE: 1997-10-28
; EARLIER FILING DATE: 1997-12-09
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-181-336-17

Query Match 39.8%; Score 779; DB 4; Length 380;

Best Local Similarity 46.3%; Pred. No. 3.7e-77;
Matches 152; Conservative 56; Mismatches 110; Indels 10; Gaps 2;

QY 43 EPNNYRTMHGRAVNSOLGKDYIQLKSLLOPIRIYSRASLYGPNIGRKNVIALLDGFM 102
Db 62 EPNSWKAHILIDGLQGEDEVEHLKLLQR-SAFONAKLYGPDIGQPRGKTVKLLRSFL 120
QY 103 KVAGSTVDVAVTQHCYIDGRVYKVMDFLTKRLDLSQIRKIQKVNTYTPGKKIWLKG 162
Db 121 KAGGEVIDSLTHWHYLYNGRIATKEDFLSSDVLDTFILSVQKILKVTREITPGKKVWLGE 180
QY 163 VYTTAGGTTNNLSDSYAGFLWNLTLGMLANOGIDVYVIRHSHFDHGYNHLYVDQNFPLPD 222
Db 181 TSSAYGGGAPLLSNTFAAGFMWLDKLGSAQMGIEVVMRQVFFGAGNHYLVNDEFLEPLD 240
QY 223 YWLSLLYKRLGPKYLAVHAGLQKRPGRVIRDKLRIYAHCTNHNHNVYRGSTTLFI 282
Db 241 YWLSLLFKLVGPRVLLSRVKGPD-----RSKRLRYLHCTNHYHPRYQEGDITLYV 291
QY 283 INLHRSRKKIKLAGTLRDKLVHQYLLQPYQGEGLSKSVQLNGQPLVMVDGTLPELAPR 342
Db 292 LNLHNVTYKHLKVPPLFRKPVDTYLLKPGPDGLLSKSVQLNGQILKWDEQTLFALTEK 351
QY 343 PLRAGRTLVIPTVMGFVFFVKNVNALAC 370
Db 352 PLPAGSALSPLAFSYGFFVIRDAKIAAC 379

RESULT 7

US-09-181-336-15
; Sequence 15, Application US/09181336A
; Patent No. 6242238
; GENERAL INFORMATION:
; APPLICANT: FREEMAN, Craig Geoffrey
; APPLICANT: PARISH, Richard
; APPLICANT: HAMDOFF, Brenton James
; APPLICANT: HULETT, Mark Darren

; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN
; FILE REFERENCE: 032505-004
; CURRENT APPLICATION NUMBER: US/09/181,336A
; CURRENT FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: AU PP0062
; EARLIER FILING DATE: 1997-10-28
; EARLIER APPLICATION NUMBER: AU PP0812
; EARLIER FILING DATE: 1997-12-09
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 532
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-181-336-15

Query Match 38.1%; Score 744.5; DB 4; Length 532;
Best Local Similarity 44.5%; Pred. No. 3.9e-73;
Matches 149; Conservative 55; Mismatches 106; Indels 25; Gaps 3;

QY 31 QFSNTYSNLI-----TEPNYRTMHGRAVNSOLGKDYIQLKSLLOPIR 75
Db 198 QWNSSNAQILLDYCSSKGYNISWELGNEPNSFLKKADIFINGSOLGEDFIQLHKLRLK-S 256
QY 76 IYSRASLYGPNIGRKNVIALLDGFMKVGSTVDVAVTQHCYIDGRVYKVMDFLTKRL 135
Db 257 TFKNAKLYGPDVGQPRKTAKMLKSLKAGGEVDSVTHWHYLYNGRATREDFLNPDVL 316
QY 136 DTLSDQIRKIQKVNTYTPGKKIWLKGVTSSAYGGGAPLLSDTFAAGFMWLDKLGLSARMG 195
Db 317 DIFISSVQKVFQVVESTREPGRKKNVIGETSSAYGGGAPLLSDTFAAGFMWLDKLGLSARMG 376
QY 196 IDVYVIRHSHFDHGYNHLYVDQNFPLDYLWLSLLYKRLGPKYLAVHAGLQKRPGRVI 255
Db 377 IEVVMRQVFFGAGNHYLVNDEFLEPLDYLWLSLLKLVGTVKLVMAVQGSKR----- 429
QY 256 RDKLRIYAHCTNHNHNVYRGSTITFIINLHRSRKKIKLAGTLRDKLVHQYLLQPYQGE 315
Db 430 --KLRYLHCTNTDPRYKEGDLTVYALNHNVTYKRLPYFSPKQVDKYLRLPLGPHG 487
QY 316 LKSKSVQLNGQPLVMVDGTLPELKPRLRAGRTL 350
Db 488 LLSKSVQLNGLTLMKVDQTLPLMEKPLRPGSSL 522

RESULT 8

US-08-435-675B-5
; Sequence 5, Application US/08435675B
; Patent No. 5710250
; GENERAL INFORMATION:
; APPLICANT: Ellis, Steven Bradley
; APPLICANT: Williams, Mark E.
; APPLICANT: Harpold, Michael Miller
; APPLICANT: Schwartz, Arnold
; APPLICANT: Brenner, Robert
; TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA: US/08/435,675B

FILING DATE: 05-MAY-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/314,083
 FILING DATE: 28-SEP-1994
 APPLICATION NUMBER: US 07/914,231
 FILING DATE: 13-JUL-1992
 APPLICATION NUMBER: US 07/603,751
 FILING DATE: 08-NOV-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Seidman, Stephanie L
 REGISTRATION NUMBER: 33,779
 REFERENCE/DOCKET NUMBER: 6362-53193
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619-238-0999
 TELEFAX: 619-238-0062
 TELEX:
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1106 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: internal
 US-08-435-675B-5

Query Match 4.5%; Score 88; DB 1; Length 1106;
Best Local Similarity 18.7%; Pred. NO.2;
Matches 68; Conservative 55; Mismatches 119; Indels 122; Gaps 14;

RESULT 9
US-08-336-257A-8
; Sequence 8, Application US/08336257A
; Patent No. 5726035
; GENERAL INFORMATION:
; APPLICANT: Jay, Scott D
; APPLICANT: Ellis, Steven B.
; APPLICANT: Harpold, Michael M.
; APPLICANT: Campbell, Kevin P.
; TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
;

ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/336.257A
FILING DATE: 07-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 54898
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 238-0999
TELEFAX: (619) 238-0062
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1106 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
US-08-336-257A-8

[illegible]

RESULT 10
US-08-455-543A-53
; Sequence 53, Application US/08455543A
; Patent No. 5792846
; GENERAL INFORMATION:
; APPLICANT: Harbold, Michael


```
; APPLICATION NUMBER: US 07/176,899
; FILING DATE: 04-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 52516 (P519739)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)238-0999
; TELEFAX: (619)238-0062
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1103 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-223-305C-53

Query Match 4.4%; Score 86; DB 2; Length 1103;
Best Local Similarity 18.7%; Pred. No. 3.3;
Matches 68; Conservative 54; Mismatches 120; Indels 122; Gaps 14;

QY 92 KNVIALLDGFMKVGSTDAVTWQHCHYIDGRVYKVMDFLTKRLDRLD-----140
Db 252 KDMILIVD-----VSGS-----VSGLTCLKLRTSVSEMLETLSDDDFNVASFNS 296
QY 141 ---QIRKIQKVNTYTPGKKIWLEGVVTTSAGTNNLSDSYAAGFLWL-----NT 187
Db 297 NQADVSCFQHLVQANVRNKKVLKDAVNNITAKGITDKGFSAPFQOLLNYNVSRANCNK 356
QY 188 LGMLANQGDVIRHSFFDHGYNHLV-----DQFNPLPDYWLSL-----227
Db 357 IIMLFDDGGEERAQAEIFKNYKDKVRFVSFGVGHYERGPLOQWACENKGYVYIPI 416
QY 228 -----LYKRLIG-PKVLA-----VHVAGLQKRPGRVIRKRLIYAHCTN 267
Db 417 GAIRINTQETLDVLGRPMVLGDKAKQVQWNTNYLDALEL-----GLVITGLTFVFNITGQ 472
QY 268 HNNHNVRSITLFIINLHRSRKKIKLAGTLRDKLVHLYLQYQGEGLSKSVQLNGQP 327
Db 473 FENKTNLKNOLILGVMGVDVSLEDIK-----RLTPRTLCPNGY-----YFAIDPNGYV 521
QY 328 LVNVDDGTLPKLPRLRAGRTLIVP-----PVTMGFEVVKNNNALA 369
Db 522 LLH-----PNLQPKPIGVG-----IPTINLKRPRPNIQNPKSOEPVTLDFLDALENDIK 571
QY 370 CRYR 373
Db 572 VEIR 575

RESULT 12
US-08-989-925-1
; Sequence 1, Application US/08989925
; Patent No. 5989820
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN ADIPOPHILIN-LIKE PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
```

```
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/989,925
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0440 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 434 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAITUT13
; CLONE: 1620223
US-08-989-925-1

Query Match 4.3%; Score 84.5; DB 2; Length 434;
Best Local Similarity 22.1%; Pred. No. 1.1;
Matches 79; Conservative 56; Mismatches 146; Indels 77; Gaps 16;

QY 7 QHFDVMELEQREKAAQMHVLLK-EQFSNTYSNLLITPN-----NYRTMHCRV 55
Db 21 QPSPV-----DRVASMPLISSTCDMVSAAYASTKESYPHVKTVCDAAEKGVRTLTA 75
QY 56 NGSQKGDYIQLKSLQPIRIYSRASLYGNIPRKNVIALLDGFMKVGSTDAVTWQ 115
Db 76 SGAQ-----PILSKLEP-QIASASEYAHRLDKLEENLFILOOPTKVLADTKELVSSK 128
QY 116 HCYIDGRVVKVMDFLTKRLDRLDLSQIRKIQKVNTYTPGKKIWLEGVVTTSAGTNNLS 175
Db 129 VSGAQEMVSAKDTVATQLSEAVDATRGAVQSGVDK-----TKSVVT---GGVQSV 177
QY 176 DSYAAGFLWNTLGMLANQGDVIRHSFFDHGYNHLVDQFNPLPDYWLSLYKRLIGP 235
Db 178 GS-----RLGQMYLGSVDTVLGKS-EEWADNHL-----PLTDAELARIATSLDGF 221
QY 236 KVLAVHVAGLQKRPGRVIR-----DKLRIYAHCTNHHNHNHNVRSITLFIINLHRSR 290
Db 222 DVASVQ-----QQREQSYFVRLGSLSERLQHAH---EHSGLKLRATKQRAQALLQSQ 274
QY 291 KIKLAGTLR-----DKLV-----HQYLL-----QPYGQEGLSKSVQLNGQPLVMVDD 333
Db 275 ALSILMETVKGVDQKLVGEGEKLHQMWLSSNNQKLOGPEKPEPPKPEQVESRALTMPRD 332

RESULT 13
US-09-175-928-4
; Sequence 4, Application US/09175928A
; Patent No. 6312921
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Mi, Sha
; APPLICANT: Genetics Institute, Inc.
```


; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-297-494-39

Query Match 4.2%; Score 82.5; DB 1; Length 921;
Best Local Similarity 24.1%; Pred. No. 6;
Matches 61; Conservative 43; Mismatches 106; Indels 43; Gaps 13;
QY 11 VMLELQRE-KAAQMHVLILKE---QFS-NTYSNLIITPNNYRTMHGRAVNGSQLG---- 61
Db : ||: ||: : ||: | | | : ||: : ||: ||
228 VLQYLQQTQASRCCLLVSEDNQLSCKVIQDKVLEEEISPLATGR-----LGQVVE 281
QY 62 -KDYIQLKSL---LQPIR---IYSRASLYGPNIGRPKNVIALLDGFMKVAG---STV 109
Db : ||||| : ||: : ||: | | | | ||: | | : |
282 DKKSILKDLTSEDWQQLQSMIGCEVQAMLCVPVISRATDQVALACAFNKLGGDLFTDQ 341
QY 110 DAVTWQHC--YIDGRVVKVMDFLKTRLLDLSQIRKIQKVNTYTPGKKIWLEGVWVTS 167
Db : ||| | : : | | : | : : : | : : | :
342 DEHVIQHCFTYSTVLTSTLAFQKEQKLKCEQALLQVAKNLFTHLDDVSVLLQELITEA 401
QY 168 AGGTNNLSDSYAAGFLWL--NTL-----GMLANQGDVVIIRHSFFDHGYNHLVDQNFN 218
Db : ||| : | | | | : | : : : | : : |
402 ----RNLSNAEICSVFLLDQNELVAKVFDGQVVEDESEYIRIPADQGIAGHVATTGQILN 457
QY 219 PLPDYWLSLLYKR 231
Db : || : | : |
458 -IPDAYAHPLFYR 469

Search completed: July 30, 2002, 08:15:09
Job time: 291 sec

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)